31.12.2019

LKB1, AMPK and AMPK-Related Kinases Regulate Fibroblast Metabolism and Mitochondrial Biogenesis Contributing to Cancer-Associated Fibroblast Transformation

PRO GRADU - 50107

Päivinen, Pekka Johan Pellervo
014330481
Genetic Bioinformatics - Genetics
Faculty of Biological and Environmental Sciences
University of Helsinki

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Tiedekunta – Fakultet – Faculty Bio- ja ympäristötieteellinen tiedekunta	Koulutusohjelma – Utbildingsprogram – Degree Programme Biologian koulutusohjelma	
Tekijä – Författare – Author Pekka Johan Pellervo Päivinen		
Työn nimi – Arbetets titel – Title LKB1, AMPK and AMPK-Related Kinases Regulate Fibroblast Metabolism and Mitochondrial Biogenesis Contributing to Cancer-Associated Fibroblast Transformation LKB1, AMPK ja AMPK-sukuiset kinaasit säätelevät fibroblastien metaboliaa ja mitokondrioiden biogeneesiä, jotka ovat yhteydessä syövän fibroblastien aktivoitumiseen		
Oppiaine/Opintosuunta – Läroämne/Studieinriktning – Subject/Study track Perinnöllisyystiede		
Työn laji – Arbetets art – Level Pro gradu	Aika – Datum – Month and year 12/2019	Sivumäärä – Sidoantal – Number of pages 137

Tiivistelmä – Referat – Abstract

Cancer-associated fibroblasts (CAF) form a heterogenous stromal cell population of a solid tumor. They are known to promote tumor growth and survival through metabolic reprogramming and inflammation. It is unclear though whether CAF are crucial component of tumor initiation and whether CAFs are dispensable altogether from the fully developed neoplasm. Tumor suppressor LKB1 regulates AMPK and AMPK-related kinases (ARK), and its function is compromised in familial disorder Peutz-Jeghers syndrome (PJS). Fibroblast specific haploinsufficiency of LKB1 alone is sufficient of initiating gastrointestinal polyposis but the mechanism through which LKB1 mediates this is only partially understood. We provide evidence that LKB1 is downregulated in multiple human malignancies including high grave serous ovarian cancer (HGSOC). Human ovarian cancer is the most lethal gynecological disease, characterized by metastasis of omentum. Loss-of LKB1 in ovarian fibroblasts was accompanied with metabolic changes associated with CAF-transformation. We screened down critical LKB1 substrates through transcriptomic and functional assays revealing AMPKa1 and MARK3 as potential downstream effectors of oxidative phosphorylation. AMPKa1, MARK1 and SIK-family were the glycolytic counterparts. We also took an initiative of cataloguing published human cancer stroma data in order to gain more comprehensive look of tumor heterogeneity. Metabolic rewiring was also observable in published cancer-stroma datasets. Human cancer-stroma divided into metabolically active and highly inflamed subtypes. These results highlights LKB1's role as a conserved metabolic caretaker in fibroblasts. Our data also support mechanistic model in which LKB1 and ARKs regulate mitochondrial metabolism, essential for CAF transformation.

Syövän mikroympäristön (strooman) fibroblastit muodostavat monimuotoisen solupopulaation. Syövän fibroblastit kykenevät edesauttamaan syövän kasvua ja selviytymistä mukauttamalla niiden metaboliaa ja aiheuttamalla tulehdusreaktioita. On kuitenkin epäselvää, kykenevätkö syövän fibroblastit itsestään aiheuttamaan syövän synnyn ja ovatko ne tarpeettomia täysin kehittyneessä syövässä. Tuumorisuppressori LKB1 säätelee AMPK:a ja AMPK-sukuisia kinaaseja ja sen toiminta on estynyt perinnöllisessä Peutz-Jegherin syndroomassa. LKB1 heterotsygoottisuus yksistään ruuansulatus järjestelmän fibroblasteissa kykenee aiheuttamaan polyyppejä mutta mekanistinen selitys on hämärän peitossa. Meidän tuloksemme osoittavat, että LKB1 tasot ovat laskeneet useissa eri syöpätyypeissä, mukaan lukien seroosissa munasarjasyövässä (high-grade serous ovarian cancer). Munasarjasyöpä on tappavin gynekologinen sairaus, ja muodostaa usein etäpesäkkeitä vatsapaidan alueelle. LKB1:n poistaminen munasarjafibroblasteista johti syövän fibroblasteja muistuttaviin metabolisiin muutoksiin. Me seuloille LKB1:n substraatit transkriptomisin ja toiminnallisten menetelmin, jotka paljastivat AMPKa1:n ja MARK3:n toimivan LKB1:n tavoin oksidatiivisessa fosforylaatiossa. AMPKa1, MARK1 ja SIK-perheen substraatit puolestaan toimivat LKB1:n tavoin glykolyysissä. Otimme myöskin aloitteen julkaistujen syöpä-strooma-tietokantojen luetteloinnista, jotta saisimme kokonaisvaltaisemman näkemyksen syöpäkudoksen monimuotoisuudesta. Metabolinen uudelleenohjelmointi oli läsnä myöskin julkaistuissa kokoelmissa. Syöpien stroomat jakautuivat metabolisesti aktiivisiin ja korkeasti tulehtuneisiin alatyyppeihin. Tuloksemme korostavat LKB1:n roolia solumetabolian ylläpitäjänä, joka on korvaamaton myös fibroblastien mitokondriaalisessa aineenvaihdunnassa. Tulokset tukevat myös mekanistista mallia, jossa LKB1 ja sen substraatit säätelevät solujen energiametaboliaa, joka puolestaan on oleellista syövän fibroblastien synnylle.

Avainsanat – Nyckelord – Keywords

Ohjaaja tai ohjaajat – Handledare – Supervisor or supervisors
Tomi Mäkelä

Säilytyspaikka – Förvaringställe – Where deposited

Muita tietoja – Övriga uppgifter – Additional information

Abstract

Cancer-associated fibroblasts (CAF) form a heterogenous stromal cell population of a solid tumor. They are known to promote tumor growth and survival through metabolic reprogramming and inflammation. It is unclear though whether CAF are crucial component of tumor initiation and whether CAFs are dispensable altogether from the fully developed neoplasm. Tumor suppressor LKB1 regulates AMPK and AMPK-related kinases (ARK), and its function is compromised in familial disorder Peutz-Jeghers syndrome (PJS). Fibroblast specific haploinsufficiency of LKB1 alone is sufficient of initiating gastrointestinal polyposis but the mechanism through which LKB1 mediates this is only partially understood. We provide evidence that LKB1 is downregulated in multiple human malignancies including high grave serous ovarian cancer (HGSOC). Human ovarian cancer is the most lethal gynecological disease, characterized by metastasis of omentum. Loss-of LKB1 in ovarian fibroblasts was accompanied with metabolic changes associated with CAF-transformation. We screened down critical LKB1 substrates through transcriptomic and functional assays revealing AMPKα1 and MARK3 as potential downstream effectors of oxidative phosphorylation. AMPKα1, MARK1 and SIK-family were the glycolytic counterparts. We also took an initiative of cataloguing published human cancer stroma data in order to gain more comprehensive look of tumor heterogeneity. Metabolic rewiring was also observable in published cancer-stroma datasets. Human cancer-stroma divided into metabolically active and highly inflamed subtypes. These results highlights LKB1's role as a conserved metabolic caretaker in fibroblasts. Our data also support mechanistic model in which LKB1 and ARKs regulate mitochondrial metabolism, essential for CAF transformation.

1. Contents

2. Abbreviations

αSMA	α-Smooth muscle actin
ATP	Adenosine triphosphate
AMP	Adenosine monophosphate
AMPK	Adenosine monophosphate kinase
ARK	AMPK-related kinase
BC	Breast cancer
CAF	Cancer-associated fibroblast
COL	Collagen
CT	Connective tissue
ECAR	Extracellular acidification rate
ECM	Extra cellular matrix
EMT	Epithelial to mesenchymal transition
FRT	Female reproductive tract
GI	Gastrointestinal tract
GSEA	Gene set enrichment analysis
HOF	Human ovarian fibroblast
iHOF	Immortalized human ovarian fibroblast
HGSOC	High-grade serous ovarian cancer
LKB1	Liver kinase beta 1
MARK1-4	Microtubule affinity regulating kinase 1-4
MEF	Mouse embryonic fibroblast
MF	Myofibroblast
MMP	Matrix metalloproteinase
MOCK	Non-manipulated control sample
MSC	Mesenchymal stem cell
mtDNA	Mitochondrial DNA
NUAK1-2	NUAK-family SNF-like kinase 1-2
OCR	Oxygen consumption rate
OXPHOS	Oxidative phosphorylation
OVCA	Ovarian cancer
PSC	Pancreatic stellate cell
PJS	Peutz-Jeghers syndrome
ROS	Reactive oxygen species
RT-PCR	Reverse transcription polymerase chain reaction
shRNA	Short/small hairpin ribonucleic acid
SIK1-3	Salt-inducible kinase 1-3
SCR	Scrambled control sample
SNRK	SNF1-related kinase
TGFB	Transforming growth factor beta

3. Review of Literature

3.1 Cancer-Associated Fibroblast

Cancer-associated fibroblasts (CAFs) account for a heterogenous stromal cell population malignant microenvironment. Though cancer evolution shapes the tumor surroundings to enhance tumor growth, the mechanisms are constrained to some extent by the homeostatic functions of fibroblasts. Thus, instead of jumping directly into the cancer, we will begin by supplying the very basic requisite data from healthy tissue.

3.1.1 Fibroblast

Fibroblasts comprise a major cell type in organism's connective tissue (CT). CT forms the structural frame of all the organs in body scheme. Fibroblasts are often found quiescent, nonproliferating and relatively sparse within the CT (Image I). They are often identified as small, spindle shaped cells expressing e.g. vimentin, PDGFRα, αSMA, FSP1 and N-Cadherin. They are major source of extracellular matrix (ECM). ECM supports anchorage, growth and migration of other cells within the tissue. Through interactions with ECM, fibroblasts exert mechanical force and tension throughout the body. This in order to preserve rigidity even with no muscle contraction present [1]. During homeostasis fibroblasts are quiescent. Upon injury, shift in balance quickly turns them into activated form. They are heterogenous group of cells with potential to transdifferentiate into multiple different stromal lineages following activation [2,3]. Fibroblast activation is highly linked to physiological and pathological conditions such as wound healing, scarring, granulation tissue formation, cancer progression, metastasis and fibrosis.

Connective tissue is of mesenchymal origin; connective tissue of a developing organism. Mesenchyme on the other hand originates from mesoderm, one of the three germ layers in developing embryo. Mesenchymal stem cells (MSC), progenitors of mesoderm and mesenchyme, give rise to different stromal cell populations of adult connective tissue (Image I, B). Fibroblasts have been shown multipotent and indistinguishable from mesenchymal (stem) cells. Even differentiated fibroblasts

show marked similarity between MSCs. For example, surface marker expression, transcriptome and methylome as MSCs [4,5].

MSCs and transdifferentiation are not the only pathways for regeneration of fibroblasts. Later in the fully developed organism generation of new fibroblasts can result from epithelial cells that transform into fibroblasts, process called epithelial-to-mesenchymal transition (EMT). Similarly, fibroblasts are able transform into epithelial cell through mesenchymal-to-epithelial transition (MET). EMT and MET are naturally occurring in development but are also induced during a variety diseases and malignancies. Fibroblasts are generated from local MCSs, but they can also migrate from bone marrow derived MCSs that spread via circulation.

3.1.2 Characteristics of Fibroblasts

In addition to localization of fibroblasts within the connective tissue, there are several molecular markers and functions that can be used to distinguish them from other cell types.

3.1.2.1 Molecular Characteristics: Extracellular Matrix

Extracellular matrix functions as a structural component of connective tissue. It is responsible for the elasticity, tension and impact resistance of CT. It also serves as docking platform for resident stromal cells and signaling molecules secreted by them. Fibroblasts are responsible for generation and modification of ECM. They are not the only ECM modifying cells, but they have their own responsibilities and tissue specific functions in it. CT is traditionally divided into loose and thick according to the fraction ECM present in the tissue.

Despite its enormous heterogeneity, ECM's structural components shear some common molecular features. Microscopic analysis of ECM shows enrichment of filamentous mesh proteins. These mesh-components are often extremely long polypeptide chains. They form fibrous protein structures by attaching to each other through e.g. conserved aminoacids repeats. The long stretches of ECM components are often modified by degradation or post-translational modifications. Components secreted by fibroblasts comprise of **collagens** (COLI, COLII, COLIII, COLIV, COLV, COLVI), **fibronectin**, **laminin**, **proteoglycans** (perlecan), **glycoproteins** (nidogen),

glycosaminoglycans, **polysaccharides** and **elastin** [6]. In order to further modify the secreted components, fibroblasts release certain modifying enzymes. Matrix metalloproteinases, such as MMP1 and MMP3, work by degrading certain ECM components, thus modifying the environment suitable for predetermined purposes [7]. Different MMPs are specific against certain types of collagens for example. This type of specificity originates from the repetitive, signature sequences in the corresponding collagen aminoacids or though post-translational modifications.

ECM of normal tissue is usually soft, with the exception to bone. The fibers are elastic and bend in response to mechanical stress. In pathologic circumstances however, the ECM is found to get stiffer and more linearized compared to homeostatis [8,9,10]. This is proposed to give e.g. cancer cells opportunities for metastasis and vascularization: degraded ECM and liberated growth factors promote vascularization.

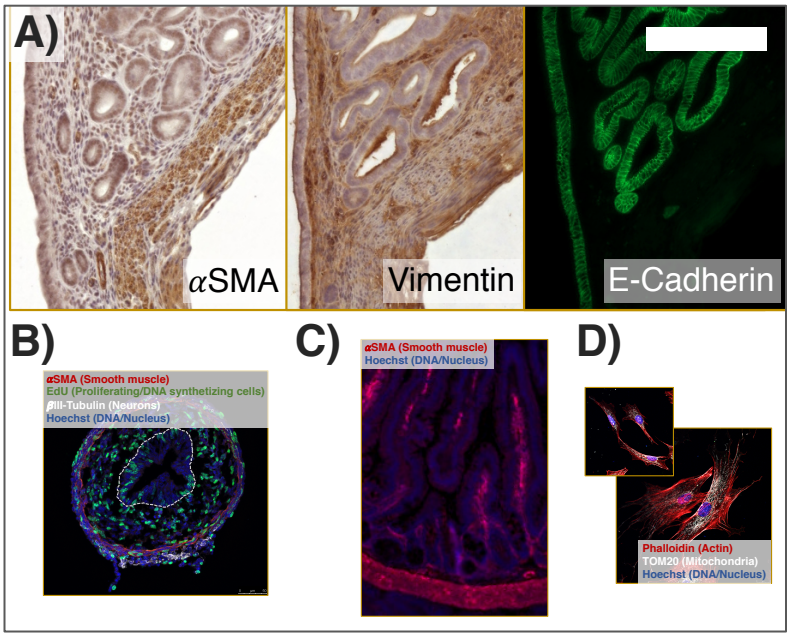


Image I. Different fibroblast populations. A) Adult murine female uterine tube: serial sections stained with αSMA (muscle layer, right), vimentin (fibroblasts, middle) and E-Cadherin (epithelia, right). Note clear separation between the αSMA+; Vim- muscle and αSMA-; Vim+ fibroblasts. B) Developing (E13.5) mouse small intestine. Mesenchymal fibroblasts locate between thin smooth muscle layer (red) and epithelial cell layer (dotted line). Highly proliferative embryonic tissue has also multiple EdU+ fibroblasts. Note, development of intestinal villi begins only after E13.5. C) Adult murine intestine villi show αSMA expressive myofibroblasts inside the villus structures in addition to smooth muscle layers (bottom). Stem cell supporting pericyptal myofibroblasts locate at the bottom crypt (not visible). D) Primary mouse embryonic fibroblasts cultured on plastic and stained with phalloidin and mitochondrial outer membrane marker Tom20. Note the size and morphological variation among the cells. C) is 2x higher magnification than A), B) and D). Scale bar 200μm. Source Päivinen P.J.P.

3.1.2.2 Functional Characteristics: Wound healing

Wound healing is one of major physiological responses highlighting the role of fibroblasts and CT. During the tissue injury, connective tissue produces granulation tissue at the site of trauma. Granulation tissue is a specialized site which supports wound closure, vascularization and immune response. It aims to keep the trauma localized and prevent it from spreading. Granulation tissue is also high in fibroblast content. Granulation tissue's environmental factors (ECM tension, secreted growth factors, cytokines and chemokines) promotes fibroblast differentiation towards so-called activated fibroblasts, myofibroblasts (MF) [11,12]. Myofibroblasts are smooth-muscle-like cells with high expression of alpha-smooth muscle actin (*ACTA2*, protein α-SMA) [1]. They are differentiated fibroblasts that promote wound contraction, scarring and fibrosis. MFs share characteristics of fibroblasts of solid tumors, cancer-associated fibroblasts.

Fully differentiated MFs sense the surrounding environment through mechano-transduction. Intracellular actin stress fiber apparatus is connected to extracellular fibronectin. This connection transmits signals from changes in the ECM structure during tissue repair and embryogenesis. Stress fibers are associated with non-muscle myosin which is one discrimination between myofibroblasts and smooth/skeletal muscle. Proto-myofibroblasts are so-called premature myofibroblasts which lack α-SMA expression [13]. As the wound heals, the proto-myofibroblasts differentiate to myofibroblasts which are capable of producing increasing amounts of tension and helps sealing the wound [11].

3.1.3 Central Stromal Pathways

In the following sections we will describe some well-established signaling pathways that have been shown to play critical roles both in normal fibroblast differentiation, connective tissue remodeling and cancer-associated-fibroblast-transformation.

3.1.3.1 Epithelial to Mesenchymal Transition

Epithelial to mesenchymal transition (EMT) is a process of transformation of epithelial cells into more mesenchymal type. The process involves multiple complicated signaling pathways, many of which converge finally into reduction of epithelial

markers and induction of mesenchymal markers. EMT is a natural process that is required for proper development and homeostasis. It is a dynamic, reversible process, and also the opposite, mesenchymal to epithelial transition (MET) is described.

Epithelial cells are characterized by apical-basal polarity, that e.g. situates different intracellular effectors accordingly within the cytosol for enhanced signal transduction. The epithelial cells harbor tight, gap and adherent junctions between the neighboring cells. Epithelial markers such as E-cadherin (CDH1), claudin (CLDN1), occludin (OCLN), desmoplakin (DSP) and plakophilin (PKP1) are epithelial proteins associated within these junctions. The *downregulation* of these protein is also used as a readout of ongoing EMT. Downregulation of epithelial markers, and subsequent induction of fibroblast markers, is accompanied by phenotypic changes. These include fibroblast-like phenotype and detachment from epithelial layer.

EMT is important for cancer invasiveness. Normal epithelial cells are not able to penetrate the basement membrane, on top of which they sat on. During EMT however, they acquire stromal features that enable them to do so. In addition to invasiveness, EMT helps cancer cell migration, supplies more stromal cells, promotes their proliferation and most importantly helps the cancer cells to adapt to the changing environment.

3.1.3.2 TGFB-signaling

Transforming growth factor beta 1 (gene *TGFB1*; protein TGFB1) is a well characterized growth factor that is capable of inducing fibroblast differentiation into myofibroblast in favorable environment [12]. In homeostasis it inhibits proliferation. Paracrine and autocrine secretion of TGFB-substrates (TGF-B1, -2 and -3) accumulate in ECM during wound healing. It is secreted from a variety of sources most notably from platelets, macrophages and epithelial cells. In order to activate TGFB gets cleaved from LTF (lactoferrin). Active form of TGFB can signal through TGFB receptors I/II (*TGFBR1/2*, TGFBR-1/2).

TGF-B1/3 homodimers bind to the TGFBR-2. This ligand-receptor complex has high affinity towards the TGFBR-1. TGFBR-1 can thus associate to TGFBR-2. Furthermore, it will bind the unoccupied TGF-B1-3 monomer. This way, ligand

binding brings the two receptor types in close proximity and transduces the extracellular signal across plasma membrane. Intracellular Ser/Thr-kinase subunit of TGFB-R2 phosphorylates the TGFB-R1 and thus activates this. The signal is further transmitted as phosphorylation series of downstream effectors: phosphorylation promotes association of CO-SMAD (SMAD4) and R-SMADs (SMAD1,2,3,5,8). CO-SMAD/R-SMAD complex migrates to the nucleus. Functioning as a transcription factors it promotes TGFB-responsive gene expression. This is mediated either directly from the membrane or via early endosome: TGFB-R1-2 complexes can be internalized, and the signalosome can be thus targeted somewhere else. Signaling can also be blocked through I-SMAD/caveolin-1 mediated endocytosis followed by proteosomal degradation. I-SMADs (SMAD6,7) directly compete with R-SMADs in TGFB-RI phosphorylation site^{74,129} [14,15]. Caveolin-1 (*CAV1*, CAV1) inhibition acts though TGFB-receptor dynamics. Caveolae are membrane bound proteins required for membrane invagination. CAV1 inhibits TGFB-signaling via promoting TGFBR-1 degradation: TGFB receptors are bound to CAV1 localized lipid-rafts that are internalized. Regulating CAV-1 expression also regulates TGFB-receptor turn-over [14]. Therefore, silencing CAV1 results in alleviated TGFB-signaling and fibroblast activation. Also, RAS is able to convey silencing phosphorylation of R-SMADs though MAPK-pathway and thus inhibiting TGFβ signaling [15].

SMADs control transcription of a range of pathways. Upregulation of *SNAI1/2*, *TWIST1, COL-VI, VIM, FN1* and downregulation of *CDH1* and *OCLN* are either direct or indirect outcomes of TGFB-signaling, but they are direct in sense of EMT [16]. TGFB can also control EMT through pathways outside SMADs like PI3K-AKT, RAS and GSK3β. It is noteworthy that TGFB-signaling works in concert with β-catenin (WNT-independent). β-catenin enables e.g. αSMA induction following TGFB activation. Loss-of β-catenin does not however compromise e.g. SMAD2 mediated transcription [17].

3.1.3.3 Integrin-signaling

Epithelial cells observe and anchor to their surroundings (bottom basement membrane and neighboring epithelial cells) through dimeric integrin molecules. Integrin dimers form from α and β subunits. Diversity of integrin subunits (18 α's and 8 β's) gives these dimeric receptors vast ability to interact with a variety of ECM components. These components include collagens, fibronectin and laminin. How epithelial integrin signaling is intertwined to fibroblasts and stroma comes through EMT. Integrin signaling works in concert with β-catenin to regulate TGFB. α3β1-integrin mediates TGFB-R1 internalization specifically at E-cadherin foci. Furthermore, this association is necessary for SMAD2 and β-catenin phosphorylation. Indeed, expression of dominant-negative β-catenin, effectively block the αSMA expression induced by TGFB1, also justifying the role of β-catenin in EMT [17].

Another integrin mediated E-cadherin disruption is mediated through β1-integrin (*ITGB1,* CD29). β1-integrin recognizes type I collagens and activates FAK (focal adhesion kinase) through Scr. FAK phosphorylates (again) β-catenin, but in contrast to earlier α3β1 induced β-catenin phosphorylation (Tyr654), this phospho-β-catenin migrates to nucleus and affects transcription (Ref[155], small notion) [18]. Thus, it is possible that these two β1-mediated integrin pathways are divergent.

3.1.3.4 Mechanotransduction: YAP/TAZ, HH and Myocardin

Transcription factors YAP and TAZ are regulated independently through conserved Hippo-signaling and direct mechanical cues from ECM and cytoskeleton. They are vital part of cellular mechanotransduction [19].

Hippo-signaling starts from MST1/2 which control the activity of LATS1/2. LATS's are responsible of regulation of Hippo pathways effectors YAP and TAZ: LATS phosphorylates YAP (Ser127) and TAZ thus inducing their cytoplasmic retention and degradation. When unphosphorylated, YAP/TAZ are transported to the nucleus. There they regulated transcription of Hippo-targets by associating with their TEAD effector proteins. Together they genes regulate proliferation, stemness and survival.

Outside-in, ECM has a profound role in inducing tension and force through the focal adhesions to the cytoskeleton, and thus promoting stiff matrix responsive genes such as α-SMA and YAP-targets. By depleting YAP from cancer-associated fibroblasts effectively blocks this mechanotransductive pathway and reduces the cancer invasiveness. Depletion of TAZ on the other hand has less pronounced effect. Importantly, also TGFB is able to induce nuclear localization of YAP though indirect ECM remodeling [20]. Additionally, this cancer supporting function of YAP can be induced by means independent of mechanotransduction, namely Hippo-signaling [20]. This could mean that despite stiff ECM, active Hippo could potentially be sufficient driver of cancer-associated fibroblast expansion.

Interesting observation from mammalian development connects YAP/TAZ, Hedgehog and myofibroblast differentiation: YAP is an inhibitor of MF differentiation [21]. It suppresses the MF gatekeeper *MYOCD* (myocardin) transcription. MYOCD belongs to a family of transcription factors responsible for myofibroblast differentiation [22]. This suppressive effect is mediated though YAP/TAZ's tissue mechanic's sensing properties and HH-signaling pathway [21].

Hedgehog signaling is a conserved signaling pathways with important implications in basic development. HH-ligand (SonicHH, IndiaHH) binding to its membrane receptor PTCH. This relieves HH-effector, Smoothened (SMO) from retainment mediated by PTCH. SMO migrates to primary cilium, were it again activates and relieves GLI transcription factors from SUFU. SUFU restricts GLIs spatially and functionally, and thus prevents HH mediated transcription. GLIs (GLI1 and GLI2) are transported to nucleus where they promote HH-responsive gene transcription. Hedgehog can contribute to cancer associated EMT. Gli1 promotes expression of Snai1 followed by reduction of E-cadherin in epithelial system and basal cell carcinoma-like expansion *in vivo*. The transcriptomic changes do not however seem to translate into morphologically mesenchymal cells, thus it is possible that HH mediates EMT only to a certain degree [23].

3.1.3.5 Cytokines, Chemokines and Inflammation

Pro-inflammatory response of the fibroblasts is a hallmark of cancer initiation and progression [24]. Fibroblast inflammatory responses are different from one cancer type to another. Skins, breast and pancreas for example show marked upregulation of inflammation whereas cervical cancer does not. The time-course of tumor inflammation can also change. The early inflammatory response enhances the macrophage recruitment, angiogenesis and survival of cancer cells. Mechanistically, inflammatory cytokines induced by the CAFs (COX-2, CXCL1, CXCL2, CXCL5, IL-6, IL-1β) could mediate more systemic inflammation of the tumor tissue through e.g. NF-κB-signaling [24].

IL-6 is one particularly interesting and well described inflammatory cytokine. IL-6 acts through binding to its membrane receptor IL6Rα. Receptor binding follows receptor dimerization and signal transduction via GP130 and STAT3 transcription factor. Forced IL-6 expression in breast cancer cells, degrades E-cadherin and induces the expression of stromal N-cadherin, SNAI, TWIST and vimentin, collectively induces EMT [25]. In human patients it is actually the stromal CAFs who are responsible for secreting IL-6. Stromal IL-6 recapitulates the EMT phenotypes but also confers significant chemoprotection to the cancer cells [26]. However, IL-6 mediated EMT seems to be mediated by other means than after loss-of α1-integrin: there the protein levels of E-cadherin remain stable but rather the localization changes [17,25,26].

3.1.4 Cancer-Associated Fibroblast

Cancer stroma comprises the micro-milieu surrounding the solid tumor. This tumor microenvironment comprises of the same stromal cells as connective tissue in general, but with tumor promoting properties. This is also the origin of **c**ancer-**a**ssociated **f**ibroblasts (CAFs). Currently the role of CAFs (carcinoma-associated-, tumor-associated-), is well acknowledged in multiple neoplasms. Generally, CAFs are taken as tumor supporting population of stromal cells. Yet, still their exact functions are heterogenous and not well understood in detail.

As in wound healing, during tumor formation, changing environmental factors directs fibroblast transformation to CAFs. Key players in transformation are the

secreted chemokines, cytokines, growth factors and other biochemically active molecules (see above) [27]. Transformed fibroblasts modify the ECM for more suitable for the new acquired state [6,28]. This can further support the activation of normal fibroblast through tissue mechanics. Since the circumstances however are not the same as in e.g. wound healing this can be taken as *misinterpretation* of the surroundings. In context of CAFs one often hears, 'cancer is a wound that does not heal'.

All organs have their own specialized connective tissues and fibroblast subpopulations [29]. Because of this, cancer can have extremely specific characteristics when it comes to initiation, progression and metastasis. To appreciate this complexity, we will go through some of best-studied organs and their malignancies: vasculature, skin, stomach and intestine, breast, pancreas and ovaries. We give explain some detailed mechanisms which apply many of the abovementioned pathways.

3.1.4.1 Vasculature

Capillary **pericytes** are proposed to possess MSC properties. They are a population of perivascular mesenchymal cells. Pericytes account for the capillary functions by communicating with vascular **endothelial cells** through paracrine signaling. This signaling accounts for capillary flow rate and permeability, trough interactions with endothelial cells. Endothelial cells form the luminal surface epithelia of blood vessels. Since all the connective tissue is also vascularized, pericytes and endothelial cells are common for all the organs. There are however tissue-specific differences in how pericytes actually colonize vessels in difference tissues at different conditions e.g. lung capillaries are more abundantly colonized by pericytes than on average. Though not fibroblasts, pericytes and endothelial cells account for disease progression through differentiation. More specifically through pericyte-fibroblast-transition (PFT) and endothelial-to-mesenchymal transition (EndoMT) respectively.

Pericytes dissociate from the vasculature after trauma or stromal cues, and subsequently acquire fibroblast-like phenotype [29,30]. PFT is augmented by PDGF-signaling: PDGF-PP binding to PDGFR-β is efficient driver for pericytes to migrate away from the vessels. This is counterbalanced by transforming the pericytes into

CAFs. Importantly, this is driven primarily through PDGFR-β and receptor type -α is not accounting for it [31]. Blocking PDGFR-β signaling, the tumor growth and angiogenesis are significantly compromised, associated with reduced PFT. Interestingly, similarly as in PFT in cancer, during spinal cord injury, resident pericytes are forming the vital part of scarring tissue. Following the injury, there is a substantial increase in vasculature followed by migration of a specific pericytes away from the vessels to form the scarring tissue. Interestingly, these scar-forming fibroblasts are positive for PDGFR-β but lose the expression of PDGFR-α. They also begin to express α-SMA and fibronectin [30]. EndoMT on the other hand, is driven through more classical EMT by TGFB-signaling. TGFB1 in the endothelial cells promotes transdifferentiation by increasing stromal gene expression and dampening the epithelial genes [32].

3.1.4.2 Skin

Skin fibroblasts (dermal fibroblasts) have been extensively characterized. So far there are at least 8 fibroblast subpopulations discovered: papillary[1] ($CD26^+$, $BLIMP1^-$, $LRIG1^+$, $SCA1^-$), reticular[2] ($DLK1^+$, $SCA1^-$), hypodermal fibroblasts[3], hair follicle dermal papilla fibroblasts[4] ($SOX2^+$, $CD133^+$), dermal sheath fibroblasts[5] ($SOX2^+$, $CD133^+$), myofibroblasts[6], preadipocytes[7] ($CD24^+$, $CD36^+$, $SCA1^+$, $ZFP423^+$) and pericytes [33,34,35]. These populations are separated from epithelial cell layer by basement membrane (BM), as are all the other mesenchymal compartments in the body. Only the fibroblasts in dermal papilla are situated in the epithelial side of BM.

During wound healing the $\alpha\text{-SMA}^+$ myofibroblasts from neighboring dermal tissue colonize the wounded area. Immune cells take care of inflammatory response initiation while fibroblasts begin to exert tension to the granulation tissue and seal the wound. Lineage tracing studies have given insight into which one of the fibroblast populations participate the most. So far it seems that reticular fibroblasts are a major contributor for wound healing, whereas FDBF's not so much. Also, preadipocytes contribute to dermal healing process [34].

Studies on dermal squamous cell carcinomas (SCC) have given important insight into skin cancer and its resident CAFs. Notch signaling has gained special attention [36,37]. In physiological circumstances, Notch is regulating e.g. keratinocyte

differentiation. Stromal deletion of *RBPJ* (also called *CBF-1* in human, *RBP-Jκ/CLS* in mouse, component of NOTCH-RBP-AP1 axis) results in CAF marker gene upregulation, dermal thinning, fibroblast senescence and development of keratinocyte tumors, type of skin SCC [36]. This effect could be effectively blocked with inhibition of AP1 and/or c-Fos/c-Jun signaling [36]. Interestingly, RBPJ also physically interacts and inhibits TP53 and thus apoptosis and senescence [37]. Thus, blocking RBPJ alone would result in CAF-target upregulation but also induced senescence and CAF exhaustion. On the other hand, inhibiting both RBPJ and TP53 would enable CAFs to circumvent TP53 mediated apoptosis and proliferate [37]. It is however possible that CAF genes regulated by loss-of RBPJ, would mediate coincidental inhibition of TP53. Indeed, FGF receptors and ligands are capable of inhibiting TP53: their expression is elevated in CAFs and more recently observed upregulated in human cancers also [37].

3.1.4.3 Lung

Pulmonary stroma comprises mostly pericytes, resident fibroblasts, myofibroblasts, mesothelial cells, lipofibroblasts, adipocytes and mesenchymal stem cells (*ABCG2*$^{+}$) [38,39]. Quantities of different fibroblast subtypes varies temporally which makes the classification challenging. Pericytes are responsible for maintaining sufficient gas exchange in extensively vascularized lungs. Resident fibroblasts localize to the interstitial compartment and differentiate to myofibroblasts [40]. Also bone marrow derived MSC supply to the pool of myofibroblasts [37].

In lung pathology, **n**on-**s**mall **c**ell **l**ung **c**ancer (**SNCLC**), interstitial lung disease and **i**diopathic **p**ulmonary **d**isease (IPD or **IPF**, **i**diopathic **p**ulmonary **f**ibrosis) are the most studied diseases with respect to surrounding stroma. IPF is a fatal lung disease in which normal tissue is replaced by increasing amounts of ECM due to stromal activation. Though not a cancer, we designate fibrotic fibroblasts of IPF also CAFs. The disease has no cure. In mouse, IPF is possible to induce by intratracheal injection of bleomycin. Following a month or-so, lesions arise, characterized by fibrotic activation. In addition to resident fibroblast activation, CAF reservoir is supplied at least through bone marrow and pericytes [39,41]. Regarding EMT, reports are either in favor or against [39,41,42]. For example, bone marrow derived CAFs (in IFP

context) do not respond to TGFB-stimulation by upregulating αSMA [41]. Then again, AECs (**a**lveolar **e**pithelial **c**ells) are either commencing or not commencing EMT *in vivo*. Explanation for this might lay within temporal dynamics in CAF transformation: signature CAF markers seem to fluctuate during the course of fibrosis, thus e.g. αSMA may not be stably expressed [39].

Lung cancer is the deadliest neoplasm, accounting up to millions of affections and deaths each year. Mutations in *K-RAS* and *STK11* are often identified, the former acting as a potent driver. NSCLC can be categorized in adenocarcinoma (ADC) and squamous carcinoma (SCC). A number of reports has studied lung cancer-stroma cross-talk with focus on metabolism. In late stage of ADC, CAFs are characterized by induced glycolytic capacity and decreased oxidative phosphorylation, consistent with the general hypothesis of **reverse Warburg effect** [43]. These changes are mediated through epithelia-stroma signaling and affects also the cancer cell's metabolism [44]. Stromal metabolism has a big role also in delineating the differences between the lung ADC and SCC. ADC-CAFs are more glycolytic compared to SCC. In SCC it is the cancer cells who are glycolytic [44]. Hexamine biosynthesis enzyme GFAT2 (gene *GFPT2*) shows high association CAF glucose uptake and EMT/TGFB-signaling in ADC, whereas *SLC2A1* and *G6PD* are more prognostic for SCC [44].

3.1.4.4 Gastrointestinal Tract

Stomach, intestine and colon stromal cells are situated in the intestinal connective tissue, lamina propria within mucosal layer. Gastrointestinal (GI) tract consists of multiple different mesenchymal cell populations: fibroblasts, mesenchymal stem cells from bone marrow, pericytes, myofibroblasts and interstitial cells of Cajal. Intestinal fibroblasts develop from mesoderm derived mesothelium but also from neural crest cells [45]. Adipocytes, epithelial and smooth muscle cells and have also been considered to supply stromal cells though EMT and dedifferentiation [46,47].

Intestinal mesenchymal cells and fibroblasts have profounding roles in tissue homeostasis. Through pathways such as WNT, BMP, HH and antigen presenting, they contribute to stem cell niche stability and immune modulatory functions of intestine [47]. For example, human and murine stromal cells express toll-like

receptors and MHC-I/-II and in this way also contribute to the innate and adaptive immune homeostasis.

There are multiple GI malignancies related to stromal impairment such as intestinal and colon cancer, intestinal polyposis, colitis-associated cancer (CAC), inflammatory bowel disease, Crohn's disease and senescence related parainflammation of intestine. Similar principles related to wound healing are applicable when intestinal epithelia get disrupted. Resident fibroblasts activate and transdifferentiate to CAFs through TGFB mediated transformation. CAFs are supplied through EMT and EndoMT. Importantly, also bone marrow (BM) derived MSC contribute to CAF supply in inflammation induced gastric cancer mouse model [48]. BM-MSC differentiate to αSMA expressing MFs, which again support the stem cells niche within the bone marrow. Upon tumorigenesis, the BM residing MFs supply to the CAFs in the tumor. TGFB and CXCR4 mediate the expansion and recruitment respectively. Eventually, BM derived CAFs constitute up to 20% of total CAFs. Interestingly, these BM-CAFs are proinflammatory (expressing IL-6, IL-33, IL1β and TNFα) compared to resident myofibroblasts [48]. This could indicate, that the actual proinflammatory hallmark is due to migratory BM-CAFs and not because of resident CAFs.

BM-CAFs could also contribute to further activating the tumor stroma. Resident fibroblasts of the intestine are responsive to IGF-1 and TNFα, the latter induced in expression of BM-CAFs. Stimulated by TNFα and IGF-1, intestinal fibroblasts induce production of procollagen α1 and TIPM-1. This TNFα induced matrix remodeling is independent on AP-1 and NF-κB, but reliant on TNFR2/STAT3 signaling [49]. Furthermore, TIMP-1 functions as a prognostic marker of colon and pancreatic cancer stroma and promotes CAF accumulation [50]. Given that no somatic alterations are not observed in at least in a subset of cancer, it is interesting to see that DNA of BM-CAFs is hypomethylated following transformation [48,51]. This could emphasize means by which CAFs' genomic DNA could be manipulated without mutational burden.

3.1.4.5 Breast

Mammary gland development is mediated through epithelia-mesenchymal cross-talk and branching morphogenesis, as are lung and GI tract. In addition, mammary glands continue developing during the organism's life and adapts to new physiological conditions after pregnancy, lactation, involution and puberty. Mature mammary gland mesenchyme comprises sub-populations of estrogen receptor alpha expressing (ERα) stromal cells. Quiescent fibroblasts can be located in inter ($CD105^-$/$CD26^+$/$TGFA^+$/$DPP\text{-}IV^+$)- and intralobular ($CD105^+$/$CD26^-$/$TGFA^-$/$DPP\text{-}IV^-$) spaces. These two cell types show different transcriptomic profiles: intralobular presenting more myofibroblastic related signature whereas interlobular fibroblasts represent more proinflammatory profile [52,53]. This type of separation is not visible in murine mammary glands however, thus showing limitations in animal models [53]. Later during embryogenesis also adipose tissue (secondary mammary mesenchyme) invades this mesenchyme. This will give rise to the mature fat pad [54]. Fat pad comprises a distinctive connective tissue due to high fraction of adipocytes in it. It also contains normal fibroblasts, endothelial cells and inflammatory cells.

Breast cancer (BC) is among the most prevalent cancers with millions affected annually. The disease holds major genetic risk. Mutations in genes such as *BRCA1*, *BRCA2*, *TP53*, *STK11* and *PTEN* are often identified. BC subtypes are divided according to increasing aggressiveness: basal-like, luminal-A (Lum-A), HER2 amplified and **t**riple **n**egative (**TN**, for estrogen, progesterone and HER2 receptors). We have learnt a great deal about CAF heterogeneity from studies about BC [55,56]. In human BC tumors, stromal MF proportion positively correlates with disease aggressiveness whereas the overall stromal size shows decreasing correlation with aggressive subtypes [55,57]. Within the stroma, any compartment (resident stromal, fat pad and bone marrow) exhibit a possible source of CAFs. Multiple hypothesis has been proposed for the causal fibroblast population. Once again, mesenchymal stem cells are suspected to represent at least one suspect [58]. Adipose tissue next to mammary gland is being investigated as a source of these MSCs though bone marrow and breast-intrinsic MSCs are equally feasible sources.

BC-CAFs are characterized by increased MF marker αSMA (independent of TGFB), induction of HIF1α-CXCL12 pathway and oxidative stress [57]. Specific

subtype of CAFs (CD10+ and GRP77+), is capable of supporting cancer stem cell (CSC) niche through secreted interleukin [56]. They also promoted the BC resistance against immunotherapy. Importantly these CAFs are not of bone marrow MSC origin but rather activated resident fibroblasts [56].

In addition to earlier specific subtype of BC-CAF, at least four different CAF subtypes are characterized in great detail, revealing their functions in tumor survival, proliferation and inflammatory response [55,56]. All four are identified in all disease states, at different proportions. Two of these CAF-types, other negative for all markers (1st), the other positive for PDGFRα and FSP1 (2nd), were more enriched in the healthy, juxtaposing tumor. The remaining two αSMA$^+$ groups, were associated with more aggressive TN disease and situated towards the tumors. The other (3rd), inflammatory type of αSMA$^+$ CAFs, are situated close to the solid tumor and attracts regulatory T-cells. They usher T-cells to differentiation thus enhancing immunosuppression. The remaining (4th) αSMA$^+$ subtype is more metabolically compromised. It is plausible to assume that BC-CAFs exhibit at least inflammatory and myofibroblastic subtypes which pose different roles in disease progression [55,56,57].

Inflammatory breast cancer (IBC) is yet another disease subtype with lower survival compared to normal BC [58]. Differences between inflammatory and non-inflammatory breast cancers have been attempted to dissect by using transcriptomic profiling. IBM shows upregulation of inflammatory pathways that is correlated with downregulation of e.g. endoplasmic reticulum stress signature. ER stress is typically considered anti-oncogenic. It would be of particular interest if the above mentioned inflammatory-αSMA$^+$ CAFs would actually play a role in IBC.

3.1.4.6 Pancreas

Pancreas can be divided into exocrine and endocrine compartments. The exocrine compartment is responsible for hormone secreting properties of pancreas and is colonized by special type of fibroblasts: pancreatic stellate cells (PSC) [59]. During homeostasis, PSC contain fat droplets and express desmin, vimentin and GFAP (glial fibrillary acidic protein) [59]. They have fibroblastic phenotype and respond to injuries as fibroblasts do: activate and turn myofibroblastic. Activation PSCs can be induced

by multiple different ways most notably inflammatory cytokines, growth factors (PDGF, FGF, TGFB/A, TNFα) and reactive oxygen species [59,60]. Also, culturing PSC (just as any fibroblasts) on plastic is sometimes enough to induce activation. Activated PSCs decrease the fat droplet content, induce expression of αSMA and PDGFRα/β, and secrete COL-I, COL-III, fibronectin and laminin [59,60,61]. Interestingly, co-culturing PSC together with different pancreatic cancer cell lines induces PCS proliferation, observation that is not a self-evident indicator of fibroblast activation [62]. Indeed, activated fibroblasts do not necessarily proliferate more than non-active counter-ones [63]. Nevertheless, the hallmarks of PSC activation can be partially reversed by culturing them in three-dimensional environment [64].

Chronic pancreatitis (CP) and pancreatic ductal adenocarcinoma (PDA) both emphasize the role of PSC activation. In acute pancreatitis, injury in pancreatic stroma or ducts results in necrosis, inflammation and PSC activation. Upon continuous injuring (e.g. reactive oxygen species from alcohol abuse), the activation might evolve into chronic pancreatitis [65]. Interestingly, even though the etiology and pathology of CP and PDA differ from one another, the stromal transcriptomics of them show greater similarity than might be expected [66].

Pancreatic cancer belongs to the malignancies of worst survival rates. Mutations in *K-RAS*, *TP53*, *p16^{INK4A}*, *SMAD4*, *STK11* and *MEN1* are often observed. PDA can be modeled in mouse with *K-Ras* and *p53* mutations conditionally in pancreas. Cancer cells are able to induce CAF transformation also in pancreas. Proximal to cancer cells, PSC induce their αSMA expression, thus differentiating into myofibroblastic type of CAF (myCAF). Acting through distance, cancer cells were able to induce generation of yet another type of CAF: αSMA^{low}, IL-6^{high} and IL-11^{high} inflammatory CAF, iCAF. Cytokines secreted by iCAFs, induced the JAK/STAT3 pathway within cancer cells, thus promoting the survival and proliferation [64]. A third distinct CAF population in murine PDA model is an antigen presenting CAF, apCAF [67]. Antigen presenting CAFs are CD74^{+}, HA-Aa^{+} and HA-Ab1^{+}, and capable of forming functional MHC-II complex. Through the MHC-II, apCAFs are able to communicate with resident T-helper cells. It will be of special interest to see if these apCAFs would pose immunomodulatory effects in PDA in similar way as their counter-CAFs in breast cancer (see above).

These differentiated states of PSCs are not static. The iCAF and apCAF phenotype can be effectively reverse by plating them from 3D-cultures into 2D which resulted in αSMA^{high}; IL-6^{low}, myCAF phenotype. Importantly, stellate cells are normal resident stromal cells in a variety of organs. It is still only the pancreatic stellate cells that have been actively studied as a potential CAF-source.

3.1.4.7 Ovary

Ovaries are part of internal female reproductive system. Ovarian stroma is highly vascularized and densely populated with fibroblast-like stromal cells. During normal ovulation cycle these fibroblasts differentiate to generate the follicle surrounding cell layers *theca interna* (androgen producing compartment) and *theca externa* [68].

Ovarian cancer is reported the deadliest gynecological malignancy. Classification of the disease is based on histology and mutation burden: high- (*TP53*, *BRCA1*, *PIK3CA*) and low-grade serous ovarian cancer (*KRAS*, *BRAF*), high (*TP53*, *BRCA1*, *PIK3CA*)- and low-grade endometroid cancer (*CTNNB1*, *PTEN*), clear cell cancer (*PTEN*, *PIK3CA*) and mucinous cancer (*KRAS*, *TP53*). The **h**igh-**g**rade **s**erous **o**varian **c**ancer (HGSOC) is the most lethal subtype. A variety of hereditary disorders such as Peutz-Jeghers syndrome and Lynch-syndrome, are known to predispose for ovarian cancer. Disease is characterized metastasis primarily to omentum and peritoneum (disease stage III/IV). Fibroblasts have been identified regulating the ovarian cancer invasion [28], metabolism [69,70], invasion [71] and survival (chemoresistance) [69].

The question of **mechanistic** relationship between CAF's metabolism and patient's immune system, has been troubling the field for a while. In ovarian cancer, CAFs actively supply glutathione (GSH) and cysteine for the cancer cells which enables them to escape platinum-based chemotherapy [69]. Cytotoxic (CD8$^+$) T-cells (T_C) regulate CAF's GSH metabolism through INFγ: INFγ induces GGT5 transcription and enzymatic activity, thus promoting the GSH turnover. Additionally, INFγ hindered the cysteine metabolism by preventing cystine transportation though CAF plasma membrane: INFγ induced JAK/STAT1 dependent transcriptional repression of cystine transporter component, xCT. Interestingly, these observations are only effective in

the fibroblasts. Without T_C, CAFs supply GSH for the cancer cells, which they again utilize in order to chelate the chemo reagents such as cisplatin [69].

CAFs are also key players in regulating cancer cell's glycolytic switch. CAFs promote cancer cell glycolytic capacity, inhibit glycogenesis and induce glycogenolysis [70]. Pathway begins from cancer cell secreted TGFB1. This induces p38α MAPK dependent flux of IL-6 and CXCL10 from CAFs, which is responded as phosphoglycomutase-1 (PGM1) phosphorylation and enhancement of glycolysis in cancer cells. Glycolysis positively correlates with metastatic potential of *in vivo* tumor models. Importantly, these results were not exclusive for ovarian cancer but rather global effects in a variety of CAF-cancer cell combinations [70].

Another important, yet uncovered CAF hallmark are micro RNAs (miRNA). miRNAs are proven important in ovarian cancer development [71]. miR-214 was discovered as a potent inducer of *CCL5*, *CCL7*, *CCL8* and *CXCL10* production and secretion. Cancer cells were found specifically primed for detecting CCL5 in their surroundings: it effectively enhances cancer cell survival and invasiveness.

Ovarian cancer cells are also able to promote transdifferentiation of not only fibroblasts, but also MSCs, stromal population present in HGSOC's metastatic site, omentum [63]. HOXA9 expression in cancer cells induces expression and secretion of TGFB2. This in effect, induces CAF transformation and upregulation of several mentioned CAF marker genes in resident fibroblasts and MSCs: *IL-6*, *CXCL12*, *VEGFA*, α*SMA*, *FAP*, *TGFB1* and *TGFB2* [63]. Importantly, cancer cells themselves do not supply to the CAF number through autocrine, TGFB2-mediated EMT.

3.1.4.8 Cancer-Associated Stroma

Just as a notion: CAFs represent the majority of stroma in the tumor. It is still important not to omit other stromal cell types. Also, the stromal cells might change dynamically depending on the disease stage, therapy phase and disease type.

Fibroblasts are known to be able to differentiate into adipocytes. In the context of breast cancer and mammary gland as a whole, new population of differentiated adipocytes has emerged. Breast cancer cells can initiate adipocyte differentiation driven through Wnt-signaling. This results firstly to generation of cancer-associated adipocytes and later on adipocyte-derived fibroblasts: adipocyte population

characterized by smaller lipid droplets and gain of (activated-)fibroblast markers such as FSP1, αSMA and morphology [72,73,74]. It is highly probable that this type of transdifferentiation of adipocytes, due to proximity of cancer cells, may also play a role in other cancers. Especially when there is prominent resource of adipocytes. As there is in the omentum for example, area associated with ovarian cancer metastasis.

3.1.4.9 Pro-or-Anti-tumorigenic CAFs?

CAFs have been shown beneficial for cancer cells in many respects. Still, one cannot kill the cancer just by removing them. In pancreatic ductal adenocarcinoma, depletion of αSMA-positive myofibroblasts after tumor initiation significantly decreases the overall survival [75]. MF-depletion affects most notably in ECM-remodeling, necrosis and cancer stemness. Interestingly, also vascularization is found reduced in these CAF-depleted tumors. Decreased survival is also associated with reduced number of effector T cells and increase in regulatory T cells. This immunocompromised disease can be rescued to some extent by inducing immunoresponse with anti-CTLA4 antibody [75].

CAF's role as a tumor initiating component is also slightly controversial. Performing xenograft experiments, it is generally accepted that fibroblasts (NFs or CAFs) are not able to induce tumorigenesis by themselves. Rather cancer cells together with fibroblasts are able to do so. Then again, stromal specific deletion of genes such as TGFB-R2 or LKB1 in stromal fibroblasts is potent inducer of neoplasia of prostate, forestomach, intestine and endometrium [76,77,78,79].

These experiments emphasize that CAFs might actually pose a dual role in tumor evolution: **I)** they have causal role in cancer initiation but **II)** they hinder cancer progression once it has initiated. The above experiments are however too drastic to recapitulate the effects in wider perspective. CAFs in breast and ovarian cancer are reported not to encounter genetic alterations during tumorigenesis [51]. In support of this idea is the above-mentioned model of myCAF-iCAF in PDAC: rather than committing to terminally differentiated CAF, activated PSCs are able to transdifferentiate dynamically. That is why complete depletion of them would not resemble the effects that would be met if CAFs were simply transdifferentiated back to normal fibroblasts. Surely CAF-transformation is accompanied with a range of

changes in the CT, but by trying to block this differentiation process could possibly repress or even reverse the effect. Tissue requires the resident fibroblasts for many reasons thus it is not surprising to see cancer-*ignition* after removal of them. It is also possible that cancer cells promote epigenetic changes in the CAFs that would permanently change fibroblast genome.

3.2 Mitochondria and Bioenergetics

Mitochondria are membranous organelles evolved after symbiotic events prior to engulfment of oxygen utilizing bacterial by eukaryotic cell. They function as energy producing units of cell. Utilizing localized membrane potential ($\Delta\Psi$) differences and charge carrier proteins, mitochondrial membrane enzymes catalyzes production of high-energy content intermediates. The $\Delta\Psi$ is stored as chemical energy within the high-energy bonds of molecules such as NADPH, NADH, FADH, ATP (adenosine triphosphate) and GTP (guanosine triphosphate). By enzymatic reactions these bonds can be broken into smaller, lower energy states and the free energy released can be utilized in downstream reactions. Mitochondrial structure is tightly linked to its functions. This is why mitochondrial dynamics are also interconnected to pathologies associated with the mitochondrial metabolism. Alzheimer's disease, Parkinson's disease, Charcot-Marie-Tooth neuropathy and type-2 diabetes are just a fraction of diseases characterized by mitochondrial dysfunction.

3.2.1 Mitochondrial dynamics

Taken the origin of mitochondria, one may not be surprised by how mitochondrial turn-over is regulated by dynamic fission and fusion events; similar processes as cell division and membrane fusion in normal eukaryotes and prokaryotes. Mitochondria are double membranous organelles, consisting of outer and inner mitochondrial membranes (OMM, IMM respectively). The space left between the two membranes is called **matrix** and space within IMM is called **inter-/inner membrane space**. Mitochondria have their own (maternally derived) mitochondrial DNA (mtDNA), replicated by DNA polymerase-γ. During homeostasis mitochondria are fused and fractionated according to cell's needs. Long stretching strands of mitochondria (increased fusion) supports the needs for high energy metabolism whereas more

dispersed and grainy pattern of mitochondria is typical during cell division to facilitate equal division of mitochondrial mass and mtDNA between the daughter cells.

Mitochondrial fusion is a process in which outer and inner membranes of different mitochondria fuse together. OMM fusion is mediated by MITOFUSIN-1 and -2 (MFN1/-2) and IMM fusion by optic ATROPY-1 (OPA-1) [80]. They are all GTPases, catalyzing the reaction: GTP → GDP + P. OMM fusion is mediated by the alignment of mitofusins' heptad regions (HR2, HR1 not participating to this) from opposing membranes. This will form coiled-coils between the mitofusins and tether the membranes before the actual fusion [81]. IMM fusion is acquired through OPA-1 and more specifically its proteolytically cleaved long and short isoforms. Activity of the isoforms is mediated by co-expression, mitochondrial membrane potential and Yme1L [82,83].

Fission of the membranes is mediated by dynamin related protein 1 (DRP-1) and close association of mitochondria and endoplasmic reticulum (ER). Inverted formin-2 (INF-2) works as an initiator of this fission. INF-2 is recruited to the ER once it is prenylated. From here it can initiate actin polymerization and constriction of mitochondria. Once the constriction reaches DRP-1 threshold this will take over and continue up to its dynamic range [84]. Fission is complete by dynamin 2 (Dyn-2) [85]. DRP-1 itself is recruited to the mitochondrial membrane by FIS1, MFF and MiD49/-51.

3.2.2 Citric Acid Cycle

Mitochondria are specialized in producing energy in aerobic conditions. Aerobic energy production is far more efficient compared to anaerobic glycolysis and fermentation. Cell is capable of producing 2 molecules of ATP through glycolysis whereas the oxidative phosphorylation (aerobic) yield 32-36 molecules. Molecular oxygen forms the basis of mitochondrial reactions such as oxidative phosphorylation (OXPHOS), electron transport chain and beta-oxidation.

In **citric acid cycle** (tricarboxylic acid cycle, TCA) group of metabolic enzymes work in sequence passing the previous end-product as a source material for the next enzyme. These enzymes are located mitochondrial matrix. The overall starting material is acetyl-coenzyme-A (Acetyl-CoA) derived from pyruvate, end product of

glycolysis for example. The sequence 9 enzymatic reactions work in order to oxidize the Acetyl-CoA and reducing the high-energy cofactor $NADP^+$ (nicotinamide adenine nucleotide phosphate) into its NADPH+H form. The cycle also produces reduced forms of $FADH_2$ (flavin adenine dinucleotide) and GTP (guanosine triphosphate), which will further improve the energetic efficiency of TCA.

3.2.3 Electron Transport Chain and Oxidative Phosphorylation

Reduced NADPH can act as an electron donor for the **electron transport chain enzymes** (ETC). These enzymes are located in the IMM. The electrons from NADPH are passed through a sequence of redox pairs situated as prosthetic groups (covalently joined cofactors) in the core **electron transport enzymes** (mostly different types of cytochromes). These enzymes work in such a way that the electrons flow according to decreasing electronegativity from donor to acceptor. This is accompanied by transport of protons (H+) across the mitochondrial membrane into the intermembranes space. This is what generates $\Delta\Psi$, electrochemical potential (pH) difference across the membrane. This chemiosmotic potential is harnessed for driving force for ATP production. There the biologically relevant energy is stored as a high energy bond between the γ- and β-phosphates as P-O-P bond.

ETC begins with passage of electrons form NADHP to the ubiquinone located in the mitochondrial **complex I**, and thereby reducing it. Complex I is able to act as an proton pump transferring 4 protons across the intermembrane to the matrix. In **Complex III** electrons are transferred from earlier ubiquinone to the cytochrome C. From the cyt-C electrons are used to further reducing steps eventually leading to additional 4 protons crossing the membrane. **Complex IV** transfers electrons from cyt-C to molecular oxygen, generating water and transferring more protons across the membrane. **Complex II** adds up electrons to the overall pool of ubiquinone/ubiquinol in Complex I, but does not contain proton pump properties as the other complexes.

Oxidative phosphorylation is a process where potential energy generated as an electro-chemical gradient during electron transport chain is harnessed for ATP synthesis. **ATP synthase** is a membrane protein by which the proton flow from the intermembrane space is channeled through to the mitochondrial matrix. As the

protons move they create proton-motive force that is used as free energy for the generation of the γ- and β-phosphate bond in ATP.

3.2.4 Beta-oxidation

The third mitochondrial metabolic pathway, beta-oxidation, yields also Acetyl-CoA which can be further supplied to TCA and electron transport chain. Beta-oxidation is a catabolic pathway for fatty-acid oxidation and breakdown. Beta-oxidation performing enzymes are located in the inner mitochondrial membranes. The pathway begins with translocations of cytosolic free fatty acids in to the mitochondrial matrix (inside the inner membranes). This is dependent of carnitine shuttle. Long chain fatty acids are first converted into fatty Acyl-CoA. Acyl-CoA then reacts with carnitine palmitoyltranferase (I/II) which is transported into the cristae of mitochondria. A second round of carnitine palmitoyltranferase reactions occur in order to bring the Acyl-CoA into the matrix.

Now the beta-oxidation can begin. During a series of reducing steps the fatty acid (Acyl-CoA) gets cleaved by 2-carbons, generating a molecule of Acetyl-CoA and reduced cofactors $FADH_2$ and NADH+H. There is a slight difference in the reaction chain when an odd numbered Acyl-chain enters the beta-oxidation, in comparison to even-numbered chain. The reaction chain chews Acetyl-CoA pieces of the body of fatty acid as long as it is possible (in even numbered this will go all the way through the structure). But in odd-numbered the last product of the reactions, propionyl-CoA is transformed into succinyl-CoA which can then feed the TCA.

The pathway is also able to handle both unsaturated and saturated fatty acids. With unsaturated starting material pathway includes specific steps in order to transfer the fatty acid into more suitable stereoisomer that is suitable for cleavage. cis-$\Delta^{3/4}$ bonds within fatty acid would cause problems with the formation of trans-Δ^2-bond. This is overcome by the usage of 2 additional enzymes enoyl CoA isomerase and 2,4 Dienoyl CoA reductase. These enzymes help to modify the bonding conformation in such a way that beta-oxidation can proceed. Beta-oxidation is also performed in peroxisomes.

3.2.5 Reactive Oxygen Species

Reactive oxygen species are naturally occurring cellular components, originating from pretty much all oxygen consuming events: mitochondrial respiration, hydrolysis, oxidation/reduction reactions and peroxisomal degradation:

$$O_2 + e^- \rightarrow O_2^{\bullet -} \rightarrow H_2O_2 + e^- \rightarrow {}^{\bullet}OH + e^- \rightarrow H_2O$$

The above reaction shows three commonly produced ROS compound: superoxide radical (2nd), hydrogen peroxide (middle) and hydroxyl radical (4th). Depending on the cellular pH, the protonation state and reactivity of ROS can change. Cellular ROS and free radical especially are extremely potent electron acceptors/oxidizers. It is thus not surprising that they can easily disrupt e.g. mitochondrial ETC. Importantly, OXPHOS itself acts as a ROS-source. ROS function also as a natural signaling molecules. Endogenous antioxidant enzymes balance the effects of ROS-mediated damage and desirable ROS-signaling. Examples of these enzymes account superoxide dismutase, glutathione peroxide and reductase and Vitamins A, C and E.

3.2.6 Mitochondria in Cancer

Cancer cells and stromal cells within tumor microenvironment form a tumor-promoting symbiotic relationship between each-other. With para- and autocrine signaling, cells aim to influence each other with the common objective of optimal growth (metastasis), given the input energy reservoir (vasculature). By fine-tuning their metabolism, cancer cells are thought to optimize this cross-talk. The seminal theory from O.H. Warburg presented how cancer cells were characterized by increased glycolysis compared to healthy cells. Due to reduced and inefficient vascularization and following hypoxia in solid tumors, increased glycolysis was seen as a way-out from metabolic crisis. However, glycolysis is far less efficient way of producing energy compared to mitochondrial oxidative phosphorylation. There are multiple hypothesis' again how and why cancer cells acquire this type of energy reprogramming. Possible explanations such as adaptation to hypoxia, mitochondrial stress and dysfunction, chemoresistance, mutational burden and metabolic intermediates as signaling molecules are all attractive and intertwining options.

Currently the so-called reverse Warburg effect has gained more support. This model proposes that cancer cells would actually possess high OXPHOS levels. This is augmented by intermediates supplied by the activated stroma, which on the other hand is characterized by enhanced glucose uptake, high glycolytic levels and low OXPHOS; normal Warburg one might say [86,87]. Glycolytic CAFs are able to direct their metabolism towards production of metabolites, such as lactate, glutamine (Gln) and glutamate (Glu), which they can feed to cancer cells [87]. Interestingly, the lactate-shuttle functions through ROS-induced mitochondrial dysfunction in both CAFs and cancer cells, whereas Gln/Glu was associated with decreased ROS-production [87,88]. Cancer cells again use lactate to fuel their own TCA [88].

The molecular mechanisms of this type of *dual-Warburg* are slowly being unraveled. Cancer cell are able to reduce IDH3α expression in adjacent fibroblasts through PDGF and TGFB mediated CAF-transformation [86]. More specifically, IDH3α reduction is achieved through micro RNA-424 mediated silencing of IDH3α. Reduction in IDH3α is countered by induction of HIF-1α and observed as increase in glycolytic capacity [86].

Another reprogramming mechanism works in breast cancer but also through micro-RNAs. Cancer cells produce **extracellular vesicles** (EV) packed with another type of micro RNA, miR-105 [89]. These vesicles delivered to the adjacent CAFs which respond by changing towards glycolytic metabolism. This increases intracellular levels of lactate, acetate, glutamate, pyruvate and aminoacids. Interestingly, miR-105-CAFs show dynamic adjustment of their metabolism according to the available nutrients. For example, during starvation miR-105-CAFs are able to detoxify the tissue from ammonium by metabolizing it to Glu and Gln. Changes following miR-105 are not associated with changes in fatty acid metabolism [89].

As mentioned above, mammary gland is evidently special tissue due to adjacent connective tissue with fat pad. Fat pad functions as an enriched source of energy for the surrounding cells. Through lipolysis, the adipocytes can secrete free fatty acids into the surroundings which the neighboring cells can utilize for beta-oxidation and oxidative phosphorylation. In breast cancer, cancer cells take advantage of this pathway by reprogramming the adipocytes into cancer-associated adipocytes (CAA). Production of FFAs by CAAs act as a source of nutrition for OXPHOS^{HIGH} cancer

cells. This has been also reported to induce fat depletion of adipocytes and transdifferentiation of them into CAFs. Interestingly, shear mitochondrial uncoupling with FCCP was shown to decrease lipid content of 3T3-L1 adipocytes [90]. Though the study lacked evidence to prove direct transdifferentiation of adipocytes it showed how mild mitochondrial membrane potential manipulation had effect of lipid synthesis and beta-oxidation related genes such as *Scd1/2*, *DHAPAT*, *CPT-2* and *MCAD*. It also highlights that though β-oxidation is not necessarily established in CAFs it does play a role in the cancer stroma.

Cancer stroma is stiffened during the course of tumor evolution but could tissue rigidity solely mediate metabolic changes. Indeed, growing cancer cells and CAFs on rigid substrate increases levels of glycolysis and OXPHOS [91]. Stiffening of ECM is associated with reciprocal release and uptake of Gln, Glu and aspartate (Asp) among CAFs and cancer cells. These aminoacids drive biosynthetic pathways for nucleotides and glutathione and are driven by YAP/TAZ mechanotransduction [91]. It is not however known how this type of matrix stiffening would affect normal cells. Could the rigidity be causal for CAF- or cancer cell transformation?

These metabolic alterations also provide important opportunities for intervention. Cancer cells could be for example targeted as $OXPHOS^{HIGH}$ cell population. On the other hand, once the metabolome of tumor is well enough established it could be possible to direct it though limiting nutrients or engineered metabolites. Could the CAF transformation and cancer cell metabolism be reverted by somehow softening the tissue?

3.3 LKB1 and AMPK-related protein kinases

LKB1 is a serine/threonine kinase encoded by human gene *STK11*. It is a tumor suppressor, mutated in Peutz-Jeghers syndrome (PJS) [92]. LKB1 functions as key regulator of development, stemness, metabolism and homeostasis in a variety of tissue and cell types and pathological and physiological conditions [93,94,95,96,97]. Together with coactivators STRADα/β and MO25α/β, LKB1 forms a catalytically active complex that is capable of phosphorylating and thus activating its substrates [98]. LKB1 mediates its cellular functions through its 14 downstream kinase substrates: AMPKα1/2, MARK1-4, NUAK1-2, SIK1-3, SNRK and BRSK1-2

[98,99,100,101]. These kinases belong AMPK-related protein kinases (ARK) family. ARKs share specific T-loop within their amino acid sequence. LKB1 is responsible for phosphorylating each substrate on a specific threonine residue within the T-loop, thus inducing the activation [98]. The association of LKB1 and STRAD25α is critical for kinase activities of substrates [98]. Each substrate again has different roles in different signaling pathways, in different organs at different developmental time points.

LKB1 functions as a key gatekeeper in signaling pathways such as mTOR, energy stress, apoptosis, TP53 and TGFB. For example, through AMPK-phosphorylation and sequential TSC2-activation, LKB1 mediates its inhibitory effect on mTOR and growth [102]. On the other hand, LKB1 physically associates with TP53, induces its phosphorylation and enable apoptosis [100,103]. LKB1 is a TGFB-pathway inhibitor. In epithelial cells, LKB1 inhibits TGFB-signaling and EMT in cooperation with LKB1-interacting protein 1 (LIP1) [104]. Also, in the mouse embryonic fibroblasts and PJS, loss-of LKB1 results in marked reduction in TGFB1 mRNA, secreted TGFB and TGFB-signaling, measured as MF differentiation and SMAD2/3 transcription [105,106].

3.3.1 LKB1 in Peutz-Jeghers Syndrome

The tumor suppressive function of LKB1 (liver kinase B1) was first discovered Peutz-Jeghers syndrome [92]. PJS is dominantly inherited autosomal syndrome which is manifested as hamartomatous polyposis of gastrointestinal tract. PJS is also linked to other malignancies such as breast, colon, small intestine, pancreatic, ovarian, cervical and lung cancer [107,108].

Of all of the LKB1-related pathologies, the PJS stands out in the perspective of cancer-associated fibroblasts and tumor stroma. Stromal specific deletions of LKB1 in the smooth muscle cells, mesenchymal progenitor cells, fibroblasts or even in T-cells are efficient drivers or polyposis [77,105,109]. Stromal deletion of LKB1 results in transformation of fibroblasts and concomitant expansion of the neighboring epithelia. Loss-of LKB1 results in reduced stromal TGFB-signaling, which would otherwise suppress epithelial growth [77,105]. In addition to TGFB, also IL-11/STAT3 pathway is implicated: stroma secreted IL-11 results in induction

STAT3, inflammation and polyposis of the adjacent epithelia. Importantly, polyposis could be effectively blocked by **JAK1/2**-inhibitor [105]. Interestingly, deletion of heterozygosity of LKB1 in T-cells is also able to induce polyposis [109]. T-cell mediated polyposis is associated with pronounced immune cell infiltration, IL-6 expression and activation of JAK/STAT3 similarly as in fibroblast specific deletion. Once again, polyposis could be blocked though **JAK2** inhibition [109].

3.3.2 LKB1, Mitochondria & Metabolism

LKB1 has been proposed an important role in mitochondrial function and energy metabolism. The first evidence about LKB1's association with mitochondria was by observing LKB1 localizing to the mitochondria, following initial steps of apoptosis [103]. More metabolically relevant results came later: in hematopoietic stem cells (HSC) LKB1 maintains mitochondrial integrity which is associated with overall survival of the lineage. These changes are partially independent of AMPKα1/α2 axis thus mediated through other substrates [94,95,96].

Interestingly LKB1 deletion in regulatory T-cells (T_{reg}) shows similar effects. LKB1 deletion in T_{reg}s compromise the immune cell survival, which results in T_H2-dominant inflammatory disorder and severe autoimmune disorder [97,110]. These syndromes are associated with mitochondrial dysfunction and metabolic imbalance: decreases in oxidative phosphorylation capacity, mitochondrial mass, mitochondrial membrane potential, mitochondrial DNA content, dysregulated lipid and triacylglycerol metabolism. Even more, pathways regulating unsaturated fatty acids where upregulated along with lipid accumulation. These mechanisms further strengthen the association of metabolism-survival-axis, but the absolute pathways are still missing. Histamine has been suggested as a regulator of T_{reg} function. In LKB1-null T_{reg}s, histidine decarboxylase (HDC) and histamine levels were upregulated. The metabolic pathway in T_{reg}s is possibly carried out though MARK and SIK-family substrates.

Partially explaining the mitochondrial dysfunction following loss-of LKB1 is induction ROS [111]. Deletion of *Stk11* in mouse embryonic fibroblasts results in significant induction of cellular ROS – characteristic that is not recapitulated by loss-of AMPKα's. Importantly, this ROS pathways seems to be different in fibroblasts and

hematopoietic cells: loss-of LKB1 in HSCs did not induce ROS production and neither did AMPKα's.

3.3.3 LKB1 Substrates

In the following section we will describe some hallmarks of each kinase substrate family. We try to be open minded about LKB1 dependency since in many of the studies have not thoroughly evaluated this. From here onwards we denote all LKB1 substrates collectively as ARKs (AMPK-related kinases); including AMPKs.

3.3.3.1 AMPKα1 & AMPKα2

AMPK (5'-adenosine monophosphate-activated protein kinase) catalytic subunits, α1 and α2 (encoded by genes *PRKAA1* and *PRKAA2* respectively), are most reputable for their function in cellular energy sensing. Functional heterotrimeric AMPK is assembled from $\alpha_{1\text{-}2}$, $\beta_{1\text{-}2}$ and $\gamma_{1\text{-}3}$ subunits. Regulatory $\gamma_{1\text{-}3}$ subunits are responsible for sensing the AMP:ATP ration. Cellular AMP and ATP compete in binding the Bateman domain of γ-subunit, thus allosterically regulating it. When activated by binding of AMP, conformational changes expose the T-loop Thr172 in $\alpha_{1\text{-}2}$-subunits subjecting it for phosphorylation and further activation. AMPK is phosphorylated by LKB1 and CAMKK2, the latter of which is responsible for Ca^{2+} mediated AMPK activation [99,100,112]. Also, CAMKK2 mediated activation is shown independent of cellular energy status, since ADP:ATP ratios wouldn't significantly alter the activation. Thus, the LKB1 mediated and ADP:ATP initiated activation of AMPK is the pathway that aims to ensure the energy homeostasis under low energy level.

AMPK has numerous downstream effectors, but the most fundamental is the regulation of mTORC1. AMPK phosphorylates and activates TSC2 which consequently inhibits mTORC1 via RHEB-GDP/GTP switch [102]. Both LKB1 and AMPK also regulate TP53. AMPK, as LKB1, physically associates with TP53, promotes the phosphorylation of Ser15 and cellular senescence [113]. The interdependency here is not necessarily direct though. In NSCLC for example it has been shown how deletion of LKB1 or AMPKα1/2 in TP53-null background, are manifested in opposing outcomes [114]. This might suggest of divergent regulation of TP53 by LKB1 and AMPK. Interestingly, TP53 can also act upstream of AMPK. In

mouse embryonic fibroblasts, TP53 responsive gene SESTRIN-2 phosphorylates AMPK and through that, regulates mTORC1 [115].

Other metabolic effects of AMPK are mediated by inactivating phosphorylation of ACC1 at Ser79 and HMGR which inhibits the fatty acid and cholesterol synthesis and induced β-oxidation [116]. Overall, AMPK activation attempts to shot down cellular anabolic, energy consuming pathways and promote catabolic, energy producing pathways.

3.3.3.2 MARK1-4

MARK1-4 (also called PAR1C, PAR1B, PAR1A, PAR1D or collectively PAR-1/MARKs) are LKB1 substrates functioning in cell polarity acquisition and regulation of WNT-signaling. They control the microtubular dynamics via MAPs (**m**icrotubule **a**ssociated **p**rotein). Their role has been extensively studied in polarity of epithelial cell, fibroblasts and neurons, using non-mammalian model organisms.

aPKC (**a**typical **p**rotein **k**inase **C**), component of PAR-3/PAR-6/aPKC master polarity complex, has been shown to inhibits MARK2's and MARK3's kinase activities by phosphorylating their Thr595 and Thr564 respectively. Phosphorylation, in concert with recruitment of 14-3-3, re-localizes MARK away from lateral tight junctions [117,118]. As a consequence, MARK2 and MARK3 are translocated away from lateral membranes to the cytoplasm [118].

Downstream of MARKs include Dishevelled (DSH), an important WNT-signal transducer [119,120]. However, this DSH phosphorylation by MARKs isn't really required for canonical β-catenin mediated WNT-signaling, but rather for PCP-non-canonical pathway [120]. Interestingly, DSH is also required for aPKC stability [121]. This sustains aPKC activity, leading further to phosphorylation and downregulation of MARK2.

Polarity is also regulated in LKB1-dependent MARK-phosphorylation. LKB1 phosphorylates MARK2 at Thr208. Phosphorylated MARK2 further phosphorylates MAP called Tau (Ser262/Ser356) and subjects it for degradation. This inhibits tubulin polymerization and microtubular regrowth and thus regulates the polarity acquisition [122]. MARK3 again phosphorylates ARHGEF2, yet another cellular polarity related component, at Ser151 in LKB1-dependent manner [123]. ARHGEF2 phosphorylation

relieves it from CYNLT1 sequester, promotes its 14-3-3 mediated translocation, and regulates actin stress fiber and focal adhesion complex dynamics.

LKB1-MARK signaling axis is regulating also the Hippo-YAP/TAZ pathway. MARKs 1, 3 and 4 are critical in LATS1/2 mediated YAP phosphorylation and cytoplasmic retention [124]. This regulation was mediated through Scribble (SCRIB), critical Hippo-effector. Importantly, LKB1 is known to suppress YAP-mediated transcription also in LATS1/2-independent means [125]. This inhibition is independent on mechanotransduction and AMPK/mTOR-signaling.

Animal models of MARK2, MARK3 and MARK4 have homeostatic changes such as: decreased body weight, resistance for high fat diet, improved glucose tolerance, reduced brown and white adipose tissue mass and overall increased metabolism [126,127,128]. No information about the tumorigenesis or polyposis was acquired. Moreover, these association studies are not strong enough to implicate LKB1's role in these profounding traits.

3.3.3.3 NUAK1-2

NUAK1 and NUAK2 (ARK5 and SNARK (Sucrose nonfermenting AMPK-related kinase) respectively) form a pair of ARKs participating in cellular adhesion, senescence, survival and metabolism. In addition to LKB1, also AKT and NDR2 are capable of phosphorylating NUAK1 (Ser600 and Thr211 respectively) but not NUAK2 [129,130]. NDR2 phosphorylation is happening on same site as LKB1's, but in this setting (SW480 cells), LKB1 mRNA was reported low which might explain redundancy [130]. Transcriptionally NUAK2 is modulated by TGFB-SMAD2/3 and Hippo pathways [131,132], both of which are highly associated in CAF transformation.

LKB1 activated NUAK1 and NUAK2 phosphorylates MYPT1, a regulatory subunit of the myosin light chain phosphatase [133,134,135]. MYPT1 phosphorylation at Ser445, Ser472 and Ser910 is accomplished via formation of NUAK1-PP1b-MYPT1 complex. Such complex formation is still to be confirmed with NUAK2, though phosphorylation of MYPT1 is already established [135]. Phosphorylation recruits 14-3-3 to the complex which results to overall sustained phosphorylation of MLC2 and cell adhesion.

Importantly, there is strong evidence that LKB1 dependent TP53 regulation is in part accomplished through NUAK1 [136]. In epithelial cells, NUAK1 phosphorylation of TP53 (Ser15 and Ser392) resulted in cell cycle arrest in G_0/G_1, and it was reliant on prior phosphorylation of NUAK1-Thr211 by LKB1. This might partially explain the gap in LKB1-TP53-AMPK regulation. It is also worth mentioning that though NUAK2 has not yet proven to control TP53, LKB1 has about 100 times higher enzymatic activity towards NUAK2 T-loop compared to for example AMPKα1 [98]. Thus, taken the extreme similarity between NUAK1 and -2 amino acid sequence, it could very well be that this preference to NUAKs over AMPK at the level of T-loop, would conduct to the level of TP53 regulation.

NUAK1 is also able to restrain cell cycle in without TP53, through Hippo-effector LATS1 [137]. In fibroblasts, NUAK1/2 phosphorylates LATS1 in Ser464, following LKB1 mediated activation. As we discussed in mechanotransduction, Hippo-pathway is responsible for the **non-mechanical** inhibition of YAP/TAZ.

NUAK1 also restrains cell cycle in concert with MYC [135,138]. This is reliant on the intact LKB1-phosphosite, Thr211, though LKB1-dependency is not technically proven. Here, in U2OS cells, NUAK1 knock-down inhibits mitochondrial electron transport chain and overall mitochondrial membrane potential, both affecting oxidative capacity. In HeLa cells (thus LKB1 independent), this, MYC-controlled NUAK1 and NUAK2 activation is Ca^{2+} sensitive [135]. MYC drives PKCα/β (**c**onventional PKC) expression. PCKα is key player in sustaining NUAK1 and NUAK2 protein stability and activity [135]. NUAK1 and NUAK2 also regulate mTORC1 in this setting, both independently and dependently on AMPK. More specifically, Ca^{2+}-dependent AMPK activation in HeLa cells is dependent on intact NUAK1: knocking down NUAK1 results in proteosomal degradation of both α1- and β1-subunits, followed by induced mTORC1/RAPTOR activation [135].

In central nervous system, NUAK1 has been demonstrated to carry crucial functions in axon specification in LKB1 dependent manner [139]. *In* utero deletion of cortical LKB1 results in failure in ipsi-/contralateral axon branching. Knock-down of LKB1 and/or NUAK1 in isolated neurons reduces the axonal growth, initiation and stability, without affecting the dendritic tree development. The finding that the axonal defects originates from increased mitochondrial motility within axonal microtubules,

emphasizes the LKB1-(*substrate*)-mitochondria axis also in CNS. Importantly, these functions where shown independent of LKB1's brain specific substrates BRSK1/2 and also NUAK2.

3.3.3.4 SIK1-3

Salt-inducible kinases, SIKs, are of LKB1 substrates which respond cAMP signaling, participates in epigenetic modifications, gluconeogenesis, lipid metabolism, sleep behavior and 'internal clock' adjustment.

Perhaps the most established role of SIKs comes from their tight association with cAMP signaling and CREB (**c**AMP **r**esponse **e**lement **b**inding protein). SIK1 was identified as a class II histone deacetylase (HDAC) and a CREB target gene [140]. The two CREB binding sites in SIK1 promoter are occupied during cAMP induction (e.g. forskolin). Activated SIK1 is able to directly phosphorylate and inhibit other HDACs, HDAC5 and HDAC4 at Ser259 and Ser489. This again relays as reduced silencing of MEF2C transcription factors, which are vital for (smooth) muscle development [22,140,141]. However, cAMP has a dual role in regulating SIK activity. **In long-term**, increased cAMP enhances SIK transcription through CREB, but **acutely** it also induces PKA mediated inhibitory SIK-phosphorylation (SIK1-Ser577 and SIK2-Ser587) [140,141].

Also, CREB regulation of SIKs is based deeper than direct phosphorylation. TORCs (TORC1-3, CREB coactivators) have been identified as assistants of CREB. They help establishing the transcriptional machinery for CREB [142]. SIKs mediate CREB regulation also at this level. Specifically, SIK1 and SIK2 (QSK, Qin-induced kinase) phosphorylates CRTC2 (TORC2) on Ser171, subjecting it to 14-3-3 mediated cytoplasmic retention [142,143]. PKA mediated phosphorylation of SIK2-Ser587 is again able to inhibit this retention. It is noteworthy that also AMPK is capable of phosphorylating TORC2 at the same site as SIK1.

SIKs provide also a link between LKB1 and TGFB-signaling. SIK1 expression is induced following TGFB-signaling activation. In cooperation with SMAD7, SIK1 associates with TGFB-R1, thus promoting its proteosomal degradation through ubiquitylation [144]. Thus, LKB1 could inhibit TGFB by **1)** activating SIK1 directly, **2)**

directly inhibiting TGFB itself or **3)** indirectly by **3.1)** inducing TGFB which **3.2)** induces SIK-mediated TGFB-silencing.

SIKs convey homeostasis also in central nervous system. Sleeping is a conserved behavior which aims for energetic balance/replenishment and CSN recovery. SIK3 has important role in sleep dynamics regulation (non-REM sleep, sleep need, waking duration). Specifically, point mutation at Ser551 of SIK3 (so-called *Sleepy*-variant) causes increased sleep need and NREM duration [145,146]. This behavior was traced to well-established interaction of SIK3 and PKA. At SIK3, PKA mediates its phosphorylation at Thr469, Ser551 and Ser674. The middle phosphosite (Ser551) is important in PKA mediated 14-3-3 recognition and translocation. This highlight SIK3's role at more systemic level and it will be interesting to see if LKB1 plays a role in this.

A real bombshell came about ovarian cancer metastasis and SIK2's role there [147]. In this scenario, autophosphorylation of SIK2 (Ser90, Ser343 and Ser358) actually had the lion's share in overall activity of SIK2, NOT due to LKB1 nor PKA! indeed, Ser358 is also putative AKT2 and PKA phosphosite. SIK2 activation in ovarian cancer cells was actually mediated via FFA-PLC-Ca^{2+} signaling. In addition, SIK2 was reported directly phosphorylating ACC1 and p85α (Ser156, regulatory subunit of PI3K), and these were essential for lipid signaling-controlled ovarian cancer metastases.

3.3.3.5 BRSK1-2

BRSK1/2 (SAD-B/A) are AMPK-Rs specifically expressed in neurons and pancreas. Though single knock-outs of BRSKs show no observable phenotype, double-knock-outs show defects in CNS development. Ventricular enlargement, cortical thinning and axonal/dendritic misorientation as hallmarks, notably similar together with pyramidal-specific deletion of LKB1 [148]. BRSK-null neurons are unable to polarize and subsequently failed in developing functional axons or dendrites, again similar to LKB1 phenotype in CNS [149]. interestingly, LKB1-null pyramidal progenitors, fail in axon/dendritic development but not in radial migration [148]. Overall, BRSK1/2 phenotypes in the cortex appear in LKB1 dependent manner, though direct genetic evidence is still lacking [148].

However, there are clear LKB1-independent phenotypes also. Careful analysis of BRSK1/2 (double knock-out) and LKB1 ($Nestin^{Cre}$, progenitor specific knock-out) knock-animal's neuroanatomy shows differences in peripheral nervous tissues [150]. Here BRSKs cause defective development of ventral horns [150]. These developmental defects were found LKB1-independent and more similar to neurotrophin-3 deficiency related phenotypes. More-over, BRSK-specific phenotypes were linked to NT-3 dependent posttranscriptional modifications within BRSK1 [150].

Within a neuron, BRSK2 has been detected in synaptic boutons (presynaptic terminal endings), where it associates with synaptic vesicles. At functional level BRSK2 seems to have a role in conveying signal across synapse, as measured in increase of frequency in excitatory postsynaptic potentials (EPSP) after overexpression. On the contrary, the kinase-dead mutant of BRSK2 showed decreased frequency in EPSPs. Together this implicates BRSK2s role in kinase dependent synaptic vesicle fusion within the synaptic bouton [151].

In pancreas BRSK2 has a role reminiscent to other ARKs: regulation of glucose and insulin sensitivity. Pancreatic specific BRSK2 knock-out mouse show hypoinsulemia, defective insulin secretion following glucose stimulation, decreased β-cells and pancreatic islet size. Further experimentation showed that glucose pulse stimulated the transcription of BRSK2 in mTOR1 dependent manner [152]. Given such a high association of BRSKs in energy metabolism, it would be interesting to explore possible link between BKRS-metabolism and BRKS-neural regulation.

Generally, BRSKs are phosphorylated by at least LKB1, TAK1 and CDK5 [98,149]. BRSKs again phosphorylate at least Tau and RIM1. The latter represents the mechanistic link of BRSK and synaptic vesicle fusion [149,151].

3.3.3.6 SNRK

In the phylogeny of AMPK-Rs, SNRK is found to diverge way earlier from others discussed here. Functional studies of SNRK (**s**ucrose **n**on-fermenting **r**elated **k**inase) are relatively limited to date. Published reports show SNRK functioning in vasculature and adipose tissue homeostasis. Expression-wise SNRK is highly expressed in testis, placenta, vasculature, including endothelial and hematopoietic cells and adipose tissue [101,153,154].

SNRK functions as a NOTCH signaling component in zebrafish vasculature development [153]. In adipose tissue SNRK is speculated to function in similar pathways associated with other ARKs: glucose uptake, fasting and insulin sensing [154]. Macrophage derived inflammatory factors are able to restrain SNRK transcription [154]. Phosphoproteomics of SNRK-depleted adipocytes revealed hypophosporylation of mTORC1, ACC1 and multiple mitochondrial proteins, whereas components of inflammatory signals showed increased phosphorylation [154].

These scattered results imply SNRKs functions associate with other ARKs, but the LKB1-dependency remains largely unexplored.

Overall, many of the ARKs have intertwining functions. They account for LKB1-dependent and independent functions. Because of the experimental variation (mouse/human, fibroblast/epithelial, ATP/AMP, LKB1/no-LKB1 etc.), the conclusion is very difficult to wrap up. What is currently missing mostly is the evaluation of the ARK functions in functionally relevant scenario. **They are all AMP-activation related, thus it would be critical to examine their functions always in metabolic stress.**

4. Aim of the Study

Main goal of the study was to explore whether LKB1 levels are changed in human cancer stroma. Another more mechanistic aim was to see if LKB1 mediates metabolic changes in fibroblasts as in immune cells. If yes, whether these could be connected to the cancer-associated fibroblast transformation. In no, by what other means LKB1 could mediate its tumor suppressive functions. We were also interested in studying the hierarchy between the LKB1 substrates in fibroblasts. This could potentially explain something about downstream effectors in CAF-transformation and help in dissecting the further mechanisms.

5. Results

5.1 Stromal LKB1 is Reduced in Ovarian Cancer LKB1 is mutated in many human malignancies, and Peutz-Jeghers syndrome predisposes patient for other disease aside of polyposis, such as breast, lung and ovarian cancer [92,108]. Since stromal deletion of LKB1 is sufficient driver of gastrointestinal polyposis, we wanted look if LKB1 levels are changed in other neoplasms' stroma [77,105,109]. We performed an exhaustive literature search and catalogued transcriptomic data from published datasets from cancers' stroma (Appendix I). (*We catalogued additional ~200 comparisons to cover other activated stromal from a variety of human malignancies and in vitro experimentations. Not discussed further in this report.*) Data sets did not include any published polyp-transcriptomes or sarcomas. Surprisingly, only a handful of datasets showed significant changes in LKB1 levels. Out of 78 datasets, only 7 had statistically significant (adjusted p-value ≤ 0.05) changes (Fig1A). One colon and one breast cancer data had stroma with high LKB1 whereas 2 ovaria cancer, another breast, one lung and one pancreatic cancer stroma samples had LKB1 levels significantly reduced in the stroma. LKB1 levels experienced the **deepest** drop in ovarian cancer dataset (Fig1A, B). Because of this and the fact that loss-of LKB1 in mouse oviductal fibroblasts is sufficient to drive endometrial cancer, we focused our following experimentation to human ovarian stroma and female reproductive tract (FRT) [78,79].

LKB1 is observed to be homo-deleted in 1.8-4% in cases of human ovarian cancer (**cBioPortal**, ***Nature 2014*** and ***TCGA;PanCancerAtlas***) and we did not observe reduction in epithelial LKB1 levels ovarian cancer (Fig1B). In attempt to rationalize stromal LKB1's role in ovarian cancer, we performed low-through-put PCR screen on ovarian cancer stroma markers. These markers included 57 most up- and down regulated genes in GSE40595 dataset. We generated human ovarian fibroblast cell lines (iHOF) with small hairpin RNA (shRNA) mediated knock-down of LKB1 (Fig1C). In the iHOFs, we saw large changes in 26 out of 57 targets (log fold change $\geq \pm 1.5 \leq$) following LKB1 knock-down. Out of these 26 genes, 16 were directionally similar to ovarian cancer stroma. Out of all 57, 33 genes were directionally similar. This suggests us, that indeed LKB1 may play a role in CAF transformation in human

disease. Additionally, these genes represent a novel LKB1 responsive genes in human ovarian fibroblasts. From now on, all the knock-down samples are designated as protein identifiers in CAPITALS, regardless of species (e.g. knock-down of *Prkaa2* mRNA in MEFs is written AMPKα2). RT-PCR and WB targets are designated as species specific mRNA and protein identifiers respectively.

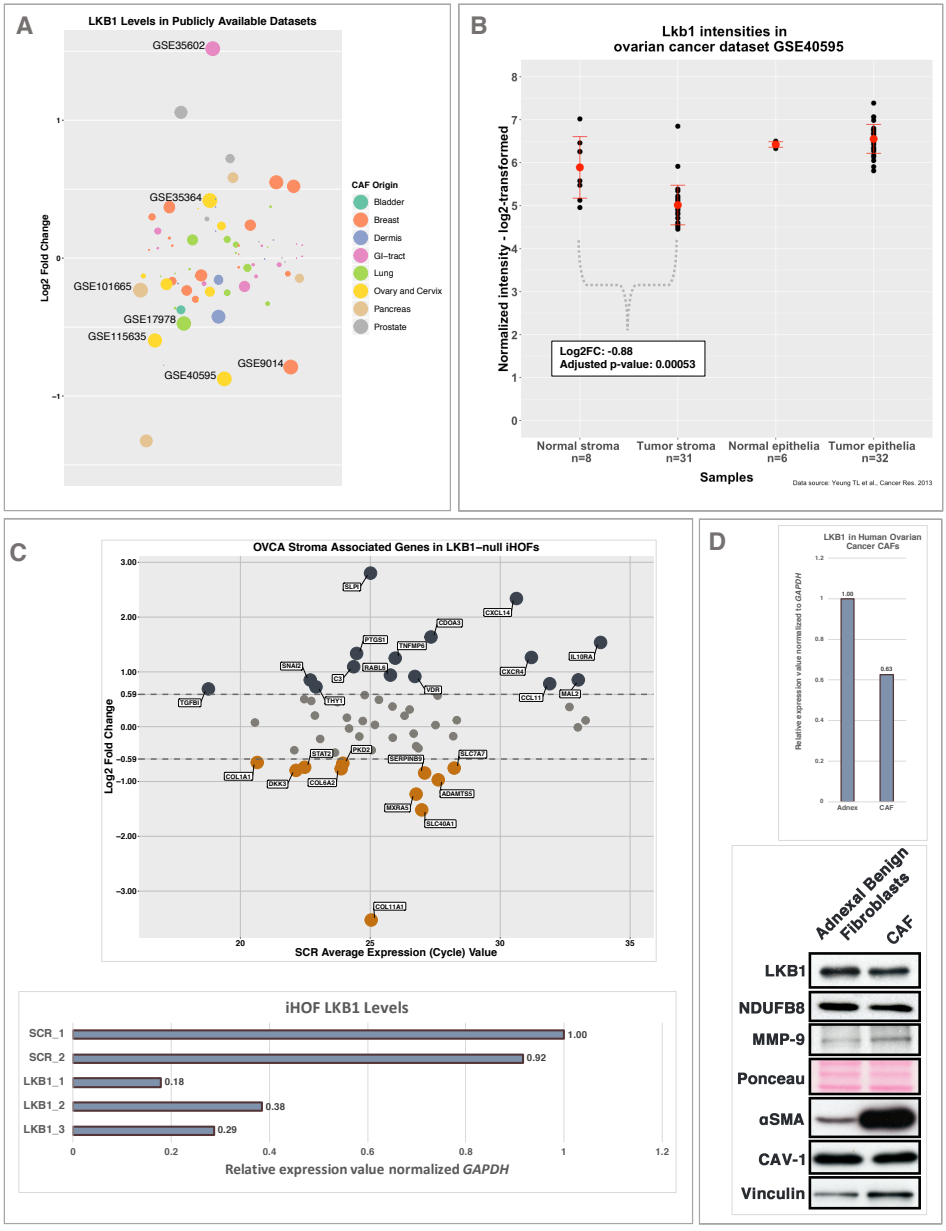


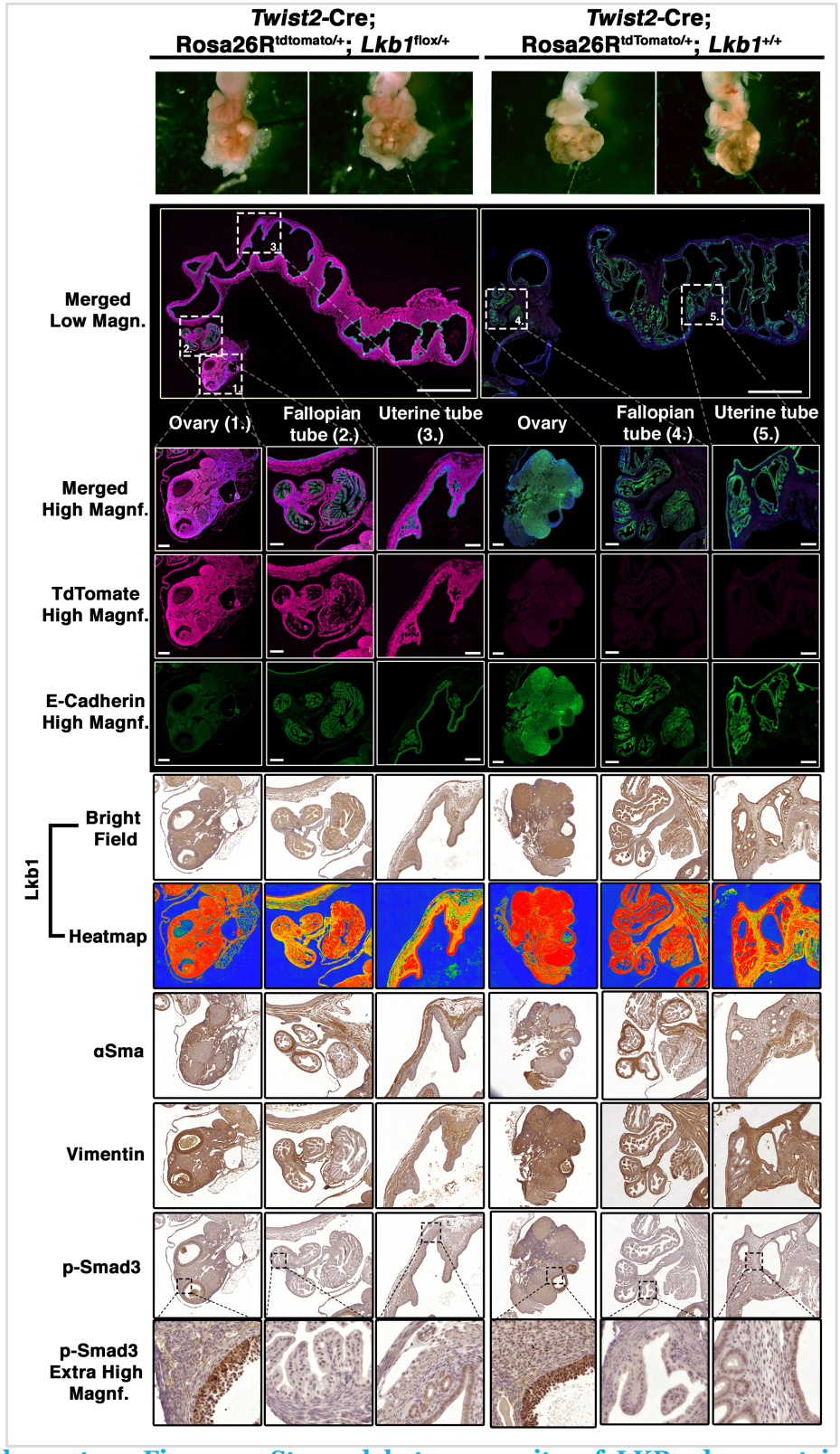
Figure 1. LKB1 levels are reduced in human ovarian cancer. A) Log-2-fold changes of 78 published cancer stroma datasets. Datasets with statistically significant changes (adjusted p-value < 0.05) are marked with their GSE-repository identifiers. B) Stromal and epithelial LKB1 mRNA levels in GSE40595. C) Low-throughput RT-PCR of human ovarian cancer stroma genes in LKB1 null iHOFs (top). Knock-down efficiencies of LKB1-iHOFs (bottom). D) LKB1 mRNA level (top) and protein level (bottom) of a single pair of human adnexal benign and ovarian cancer CAFs. What are also probed on western are, CAF-markers MMP-9, CAV-1 and αSMA. NDUFB8 was analyzed for mitochondrial biogenesis, also associated CAF-transformation. 3rd row ponceau as a loading control for LKB1, NDUFB8 and MMP-9 and bottom row Vinculin as a loading control for αSMA and CAV-1.

Since the ovarian cancer dataset (GSE40595) represented the **all** stromal cell populations, not only fibroblasts, we evaluated LKB1 levels (along with other markers) in single pair of adnexal benign fibroblasts and CAFs derived from ovarian cancer patient. We observed ~40% reduction in *STK11* mRNA and somewhat similar reduction in protein level (Fig1D). We also observed induction in CAF-markers MMP-9 and αSMA. Slight decreases were observed in mitochondrial complex I component NDUFB8 and a third CAF-marker CAV-1, markers for mitochondrial biogenesis and caveolae signaling respectively. Mitochondrial biogenesis and caveolae mediated membrane trafficking are contributing to CAF-transformation (see the literature review). These data highlight that LKB1 levels are reduced *de novo* in human ovarian cancer and that this downregulation is associated with human cancer related transcriptomic signature.

In murine model of endometrial cancer, epithelial heterozygosity of LKB1 does not drive tumorigenesis [170]. On the other hand, whole-body heterozygosity and epithelial or mesenchymal homozygous deletion of LKB1 do progress into endometrial adenocarcinoma and oviductal adenomas [78,79,170]. Stromal heterozygosity of LKB1 is known to be sufficient driver of gastrointestinal polyposis, but whether same holds true for FRT is not yet reported. Given the opportunity, we analyzed murine ovaries, fallopian tubes and uterine tubes of Peutz-Jeghers mouse model at the age of 11 months (*Twist2*$^{Cre/+}$; *Stk11*$^{+/flox}$ = *Lkb1*$^{Twist2KO/+}$). Importantly, mutant mice were fertile and by the time of collection, GI-polyposis was fully penetrant in the heterozygous animals. We did not observe noticeable difference in the size or external appearance of the ovaries, fallopian or uterine tubes between the genotypes of 2 versus 2 animals (4 versus 4 ovaries) (FigS1).

HGSOC is known to initiate through from either fallopian tubes or ovarian surface epithelia [155,156]. Thus, we analyzed these anatomical regions in closer detail. We also analyzed uterine for extra-ovarian growth. Recombination was evident in the stroma and muscle layers through-out the mouse FRT (FigS1). LKB1 downregulation was assessed by immunohistochemical analysis. αSMA was highly expressed in uterine muscle layer and fallopian tubes with no differences between the genotypes. Epithelial cell layer overlaying the tubular structures and ovarian

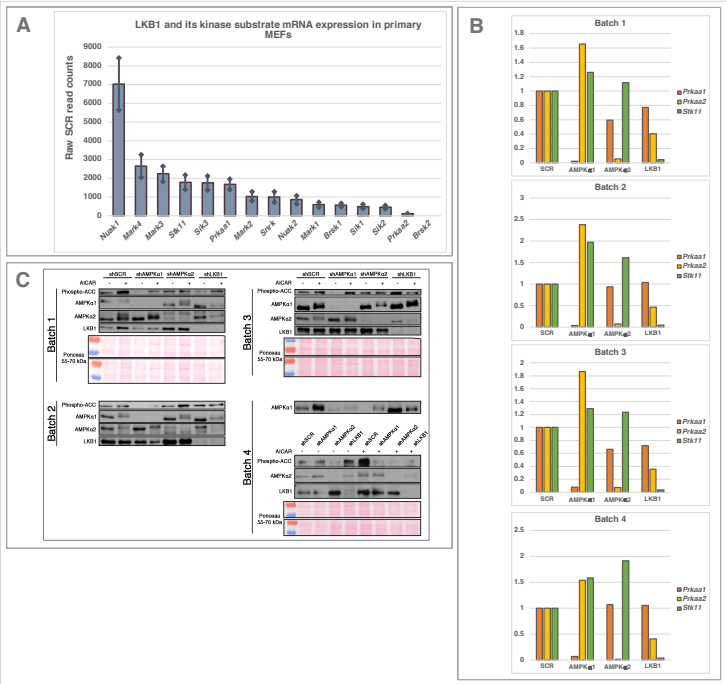
follicular structures was uniform in both genotypes. We observed significantly less E-Cadherin expression in ovarian structures of LKB1-heterozygotes which might imply ongoing EMT. To asses this, we visualized active TGFB-signaling by staining phospho-SMAD3. We did not however observe any difference: p-SMAD3 stained similarly in both genotypes, follicular epithelial layer. This ruled out active TGFB as a source of this E-Cadherin pattern. This suggested that in the FRT, LKB1 heterozygosity requires further reduction and/or concomitant oncogenic signal(s) from neighboring epithelia or stroma, in order to initiate stromal expansion and cancer progression.



Supplementary Figure 1. Stromal heterozygosity of LKB1 does not induce ovarian cancer in Peutz-Jeghers mouse model. No general differences in gross anatomy (top row) or in fine structures of heterozygous (left) and wild type (right) mouse female reproductive tracts. Recombination was assessed with TdTomato reporter (4th row) and LKB1 staining (6th and 7th rows). Wild type ovaries (Rows 3-11, column 4) were from different mouse than fallopian and uterine tube (Rows 1-11, columns 5-6). 1st row in same scale. 2nd row scale bar 2000μm. Rows 3-10 scalebar 200μm.

5.2 AMPK and AMPK-Related Kinases Contribute Significantly and Hierarchically to LKB1 Transcriptome in Fibroblasts We began dissecting the role of LKB1 in ovarian cancer evolution and CAF-transformation by evaluating though which substrates LKB1 mediates its functions in fibroblasts. LKB1 regulates 14 kinases belonging to family of AMPK-related protein kinases [98]. As a model of stromal fibroblasts, we used primary **m**ouse **e**mbryonic **f**ibroblasts (MEF). Because of inherent heterogeneity of MEF's they also provided us an opportunity to circumvent and model endogenous variability of tissue stroma [29,157]. Also, as a primary cell line, MEFs do not experience somatic modifications related to immortalization or extended culture conditions. LKB1 and 12 of its kinase substrates where silenced using short hairpin RNA (shRNA) mediated interference [158]. We did not induce knock-down of BRKS1 or BRSK2 due to their low expression and functional unrelatedness in fibroblasts (FigS2A).

Author of the thesis was responsible of **producing** sequencing batch 4: AMPKα1, AMPKα2 and LKB1 (2nd replicate). Thus, validation data is only supplied for the corresponding batch! Hairpins were validated according to knock-down efficiency (Data not shown). Knock-down efficiencies of sequencing samples were validated with RT-PCR and western blotting (FigS2B, C). Levels of AMPKα1, AMPKα2 and LKB1 reached >90% compared to SCR control in mRNA level. Importantly, we observed reproducible upregulation of AMPKα2 following AMPKα1 knock-down. This might speak for underlying hairpin effect or compensatory mechanism within AMPKs. Also, upregulation of LKB1 was observed after AMPKα1/α2 knock-downs. On the other hand, AMPKα2 levels experienced ~50% drop, following LKB1 knock-down.



Supplementary Figure 2. Validation of RNA-sequencing batch #2: AMPKα1, AMPKα2 and LKB1 (2nd replicate) samples. A) LKB1, AMPK and ARK expression in primary MEFs as represented in average read counts from SCR-control samples. B) RT-PCR analyses of knock-down efficiencies. C) Western-blot analyses of protein levels of knock-down samples. AICAR (AMPK-activator) samples were also generated though not sequenced. Ponceau stainings as a loading control: top-ponceau for AMPKα1 and bottom for p-ACC1, AMPKα2 and LKB1 (Batch 1 & 3). Note the shift in sample order in p-ACC, AMPKα2 and LKB1 in sample batch 4. No ponceau-control for Batch 2 available.

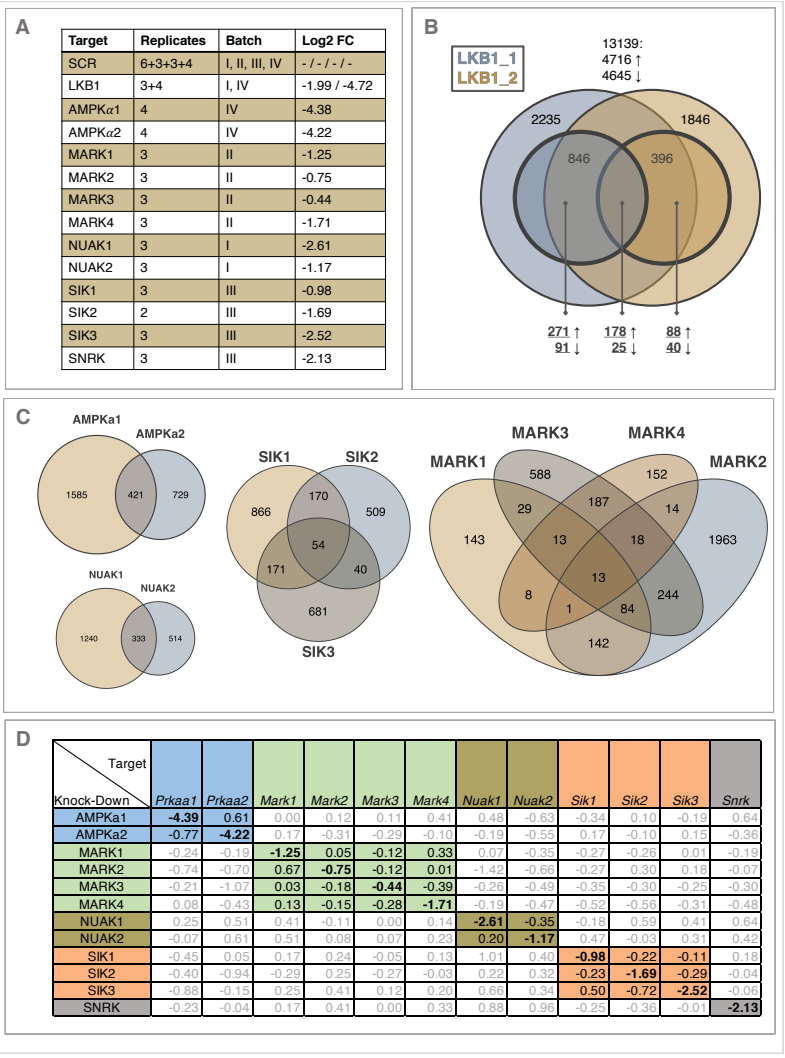
After quality check and library preparation, MEF transcriptomes were generated with single-end RNA-sequencing. Sequencing was performed in four batches (FigS3A). All batches were included with scrambled (SCR) control. Each batch included 2 to 6 biological replicates, specifically embryos. Looking into the similarity of significant genes, we observed good overlap between the two LKB1 batches (>71% directional similarity). For further analyses we combined the two LKB1 data (FigS3B). Also, without going deeper into the directionality, we could readily observe how of the ARK subfamilies related to each other (FigS3C). For example, among SIK-family, SIK2 and SIK3 were the most distant. In MARKs, MARK4 seemed the most distant showing highest amount of similarity together with MARK3. MARK2 shared plenty with MARK1 and MARK3 but it also had a huge number of exclusive genes. The absolute shared genes between AMPKα's and NUAK's were higher, though they still retained high number of exclusive genes.

We also tried to identify feed-back-loops by analyzing expression values of substrate paralogs following ARK-knock-downs (FigS3D). NUAKs, SIK1/2, MARK1/3/4 had no significant effect on each other's expressions. The previously observed AMPKα-feed-back was present also in the sequencing data, but we also saw changes in paralog expressions following MARK2 and SIK3 knock-downs: MARK2 knock-down induced expression of MARK1 and SIK3 induced SIK1 but

silenced SIK2. Whether these are true biological effects or rather hairpin-dependent will require further inspection. Interestingly, we did not observe the reported AMPKα1 reduction following NUAK1 knock-down in mRNA level [135,138].

A

Target	Replicates	Batch	Log2 FC
SCR	6+3+3+4	I, II, III, IV	- / - / - / -
LKB1	3+4	I, IV	-1.99 / -4.72
AMPKα1	4	IV	-4.38
AMPKα2	4	IV	-4.22
MARK1	3	II	-1.25
MARK2	3	II	-0.75
MARK3	3	II	-0.44
MARK4	3	II	-1.71
NUAK1	3	I	-2.61
NUAK2	3	I	-1.17
SIK1	3	III	-0.98
SIK2	2	III	-1.69
SIK3	3	III	-2.52
SNRK	3	III	-2.13



D

Knock-Down \ Target	*Prkaa1*	*Prkaa2*	*Mark1*	*Mark2*	*Mark3*	*Mark4*	*Nuak1*	*Nuak2*	*Sik1*	*Sik2*	*Sik3*	*Snrk*
AMPKa1	**-4.39**	0.61	0.00	0.12	0.11	0.41	0.48	-0.63	-0.34	0.10	-0.19	0.64
AMPKa2	-0.77	**-4.22**	0.17	-0.31	-0.29	-0.10	-0.19	-0.55	0.17	-0.10	0.15	-0.36
MARK1	-0.24	-0.19	**-1.25**	0.05	-0.12	0.33	0.07	-0.35	-0.27	-0.26	0.01	-0.19
MARK2	-0.74	-0.70	0.67	**-0.75**	-0.12	0.01	-1.42	-0.66	-0.27	0.30	0.19	-0.07
MARK3	-0.21	-1.07	0.03	-0.18	**-0.44**	-0.39	-0.26	-0.49	-0.35	-0.30	-0.25	-0.30
MARK4	0.08	-0.43	0.13	-0.15	-0.28	**-1.71**	-0.19	-0.47	-0.52	-0.56	-0.31	-0.48
NUAK1	0.25	0.51	0.41	-0.11	0.00	0.14	**-2.61**	-0.35	-0.18	0.59	0.41	0.64
NUAK2	-0.07	0.61	0.51	0.08	0.07	0.23	0.20	**-1.17**	0.47	-0.03	0.31	0.42
SIK1	-0.45	0.05	0.17	0.24	-0.05	0.13	1.01	0.40	**-0.98**	-0.22	-0.11	0.18
SIK2	-0.40	-0.94	-0.29	0.25	-0.27	-0.03	0.22	0.32	-0.23	**-1.69**	-0.29	-0.04
SIK3	-0.88	-0.15	0.25	0.41	0.12	0.20	0.66	0.34	0.50	-0.72	**-2.52**	-0.06
SNRK	-0.23	-0.04	0.17	0.41	0.00	0.33	0.88	0.96	-0.25	-0.36	-0.01	**-2.13**

Supplementary Figure 3 RNA-sequencing batch information and ARK-paralog cross-talk. A) Information about the sequencing batches. B) Aggregation of two LKB1-MEF batches. Combined LKB1-dataset included (I) significantly altered* common genes and, (II) genes which were significantly altered* in the other and directionally similar in the other. Thick circles represent the significantly altered fraction (*) and thin circles the non-significant fraction. C) Venn-diagrams representing the common significantly changed* between the ARK-paralogs. Directionality was not assessed between the paralogs. D) Table with knockdown (rows) induced changes in the other ARK's mRNA levels (columns). (*) = Adjusted p-value < 0.05 and Log-2-fold change $\leq \mp 1.5 \geq$.

After data processing, hierarchical clustering revealed LKB1 batches forming a separate branch together with previously published LKB1 polyp-data from Peutz-Jeghers mouse model (Fig2A) [105]. MARK3 and MARK4 clustered closest among the substrates to LKB1 datasets. SIKs and SNRK formed distinct cluster of four. AMPKα1, AMPKα2, MARK1, MARK2, NUAK1 and NUAK2 formed a third distinguishable group, which was closer to LKB1s than SIK-SNRK.

In order to get more detailed view on substrate contribution on LKB1 transcriptome, we analyzed gene-by-gene similarity of substrates with genes most

significantly regulated by LKB1. By applying fold change and adjusted p-value thresholds we saw 693 common genes regulated after LKB1 knock-down (FigS3B, underlined). From these MARK2, NUAK1 and AMPKα2 took the lion's share (Fig2B). Adjusting these changes to the whole transcriptomic changes, showed that AMPKα2 and NUAK2 and NUAK1 had the highest LKB1-related transcriptomic fingerprint (Fig2C). Also, looking at genes changing solely after knocking down a single substrate showed MARK2, NUAK1, AMPKα1 and AMPKα2 having the highest number of exclusive genes (Fig2D). Overall, sequenced ARKs covered ~77% of significantly altered transcriptome of LKB1.

LKB1 is known regulator of cellular pathways ranging from cell polarity, proliferation, differentiation and metabolism [93,94,103,122]. To see if our *in silico* model could capture some of these functions, we performed gene set enrichment analysis (GSEA) for LKB1 and the substrates. LKB1-MEFs showed positive enrichment of pathways such as JAK/STAT3, glycolysis and interferon responses (Fig2E). On the other hand, hallmarks such as TGFB-signaling and myogenesis were negatively enriched, which is in agreement with the function of LKB1 as a TGFB inhibitor [104]. Generally, LKB1 replicates were highly enriched in inflammatory related pathways. Looking down the substrates it seemed that more pathways were negatively than positively enriched. None of the GSEA hallmarks were shared with all the samples.

LKB1 and ARKs maintain metabolic equilibrium [94,95,96,97]. Thus, we focused our further GSEA more specifically on metabolic pathways. LKB1 replicates were differentially deregulated among oxidative phosphorylation. This was interesting in 2 ways: there is increasing evidence of decreased oxidative phosphorylation following loss-of LKB1 AND activated fibroblasts in cancer stroma are described to maintain low OXPHOS and to cancer progression [86,87,88]. Seeing this difference in OXPHOS, we conducted more detailed leading-edge analysis without being able to resolve the origin of difference (Data not shown). LKB1 replicate(s) shared positive enrichment in glycolysis, xenobiotic, and cholesterol metabolism. Thus, loss-of LKB1 results in changes in oxidative phosphorylation which could be compensated by upregulation of multiple other metabolic pathways. ARK knock-downs resulted generally in decreased glycolysis and fatty-acid metabolism. Oxidative

phosphorylation was more dispersed among them. Thus, we hypothesize that LKB1 regulates mitochondrial metabolism also in fibroblasts, and that some of its substrates possibly functions downstream of it in this pathway.

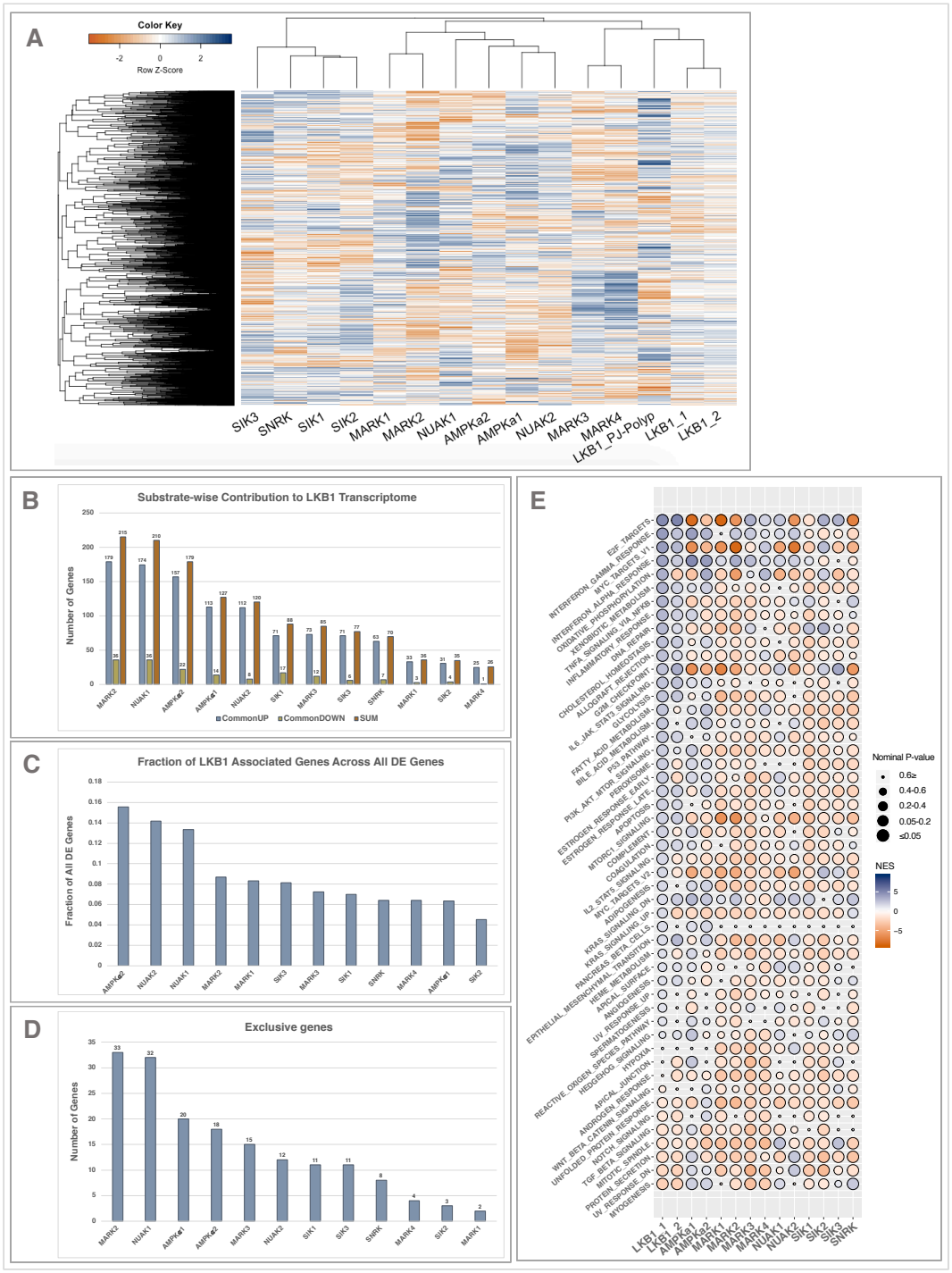
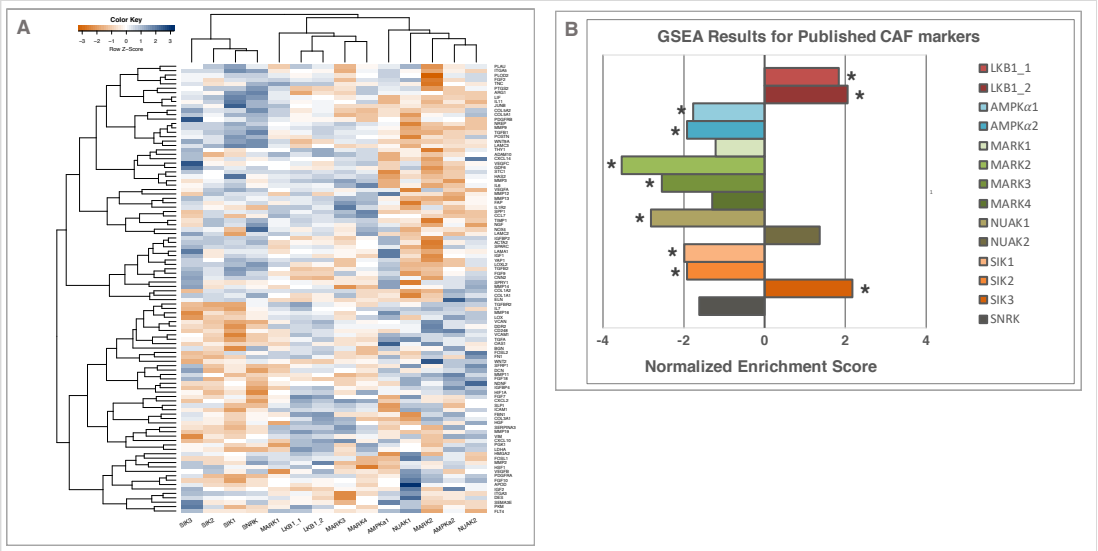


Figure 2. Sequencing analyses of MEFs with LKB1, AMPKα's and AMPK-related kinase knock-downs. A) Unbiased hierarchical clustering of LKB1-/ARK knock-down MEFs and published Peutz-Jeghers mouse model polyps. B) Directional similarity of LKB1 (FigS3B) and substrates: Up (blue), down (yellow) and total (red) among significant genes*. C) Fraction of total genes (B, red) from all the significantly differentially expressed genes within a substrate*. D) Number of exclusively, significantly expressed genes from LKB1 (FigS3B) following substrate knock-down. E) GSEA analysis of Molecular Signatures Database (MSigDB) hallmarks of LKB1-/ARK knock-down MEFs. (*) = Adjusted p-value ≤ 0.05 and Log-2-fold change ≤∓1.5≥.

5.3 Cancer-Associated Fibroblast Classification Unveils Inflammatory and Myofibroblastic CAF Subtypes Before going deeper into the mechanism driving the possible CAF-transformation, we wanted to improve our confidence about LKB1-MEFs as a good surrogate for CAFs. There is not a single all-inclusive CAF-marker that we could use as a universal identifier. CAFs are rather identified by a cluster of markers. By utilizing curated CAF-gene set [37,159], we first aimed to see how LKB1 and ARKs segregate/merge in terms of CAF-markers. By limiting the clustering to ~120 CAF-markers, did not see drastic changes in the clustering compared to initial clustering (Fig2A). MARK1 switched to LKB1-MARK3/4 cluster (FigS4A). SIK/SNRK clustered as an outgroup and AMPK-NUAK were closer to MARK-LKB1. This could indicate that SIK, AMPK-NUAK and MARK-LKB1 represents different CAF subtypes. We also performed GSEA analysis with the same CAF-gene set in order to normalize the value of these CAF-markers over to the whole transcriptome. Surprisingly, majority of the ARKs showed negative enrichment against the markers. LKB1 together with SIK3 and NUAK2 were positively enriched (FigS4B). This highlights that not all the genes within these CAF-markers are equally highly altered in all the knock-downs. Thus, the markers give a different weight for depending on the sample.



Supplementary Figure 4. Association of Curated CAF-markers in LKB1-ARK-MEFs. A) Unbiased hierarchical clustering of LKB1- and ARK-MEFs based of 166 CAF-markers. B) GSEA analysis of LKB1- and ARK-MEFs against gene set constructed from curated CAF-markers. Asterix (*) signifies statistical significance (nominal p-value ≤ 0.05).

Despite the CAF-markers are *curated*, the directionality and biological purpose of them is often unknown. Thus, we wanted to deepen our understanding in CAF-heterogeneity to the level of whole transcriptome using actual patient material. To achieve this and really benchmark our MEF-model, we took advantage of our previous literature search (Fig1A, Appendix I). We divided the samples into two distinct categories. **Direct** (29pc) samples were acquired though microdissection or FACS-sorting and were considered to represent the actual stromal signature present at the time of dissection. **Indirect** (51pc) samples were patient derived cell lines which may have lost some of their pathogenic phenotypes through prolonged *in vitro* culture conditions but included only fibroblasts. Hierarchical clustering was performed for the two groups separately which identified brute segregation of another two groups in both direct and indirect (Fig3A, B). Samples from identified clusters (Patient Derived Direct/Indirect 1 & 2 $\rightarrow$ 4 clusters) were further merged as a representative of the whole cluster.

CAFs represent heterogeneity which can roughly divide then into two fundamentally different subtypes: myofibroblast-like and inflammatory-like [64]. The former is more myogenic and metabolically active whereas the latter represents more inflamed phenotype. To see if our identified clusters could capture some of this heterogeneity, we performed GSEA analysis on the merged clusters. The direct clusters were somewhat similar, differing only in pathways such as hypoxia, oxidative phosphorylation and MYC targets (Fig3C). This was probably due to noise originating through other stromal cell types present in majority of samples. This would also suggest that the pathways that actually show difference, are very strong since they pop-out despite this noise. Thus, e.g. OXPHOS would highlight an important hallmark discriminating these two groups.

Surprisingly, indirect clusters represented more dramatic differences among each other. Cluster one was positively enriched in oxidative phosphorylation, glycolysis and myogenesis and generally negative in inflammation associated hallmarks. Indirect cluster 2 on the other hand was positively enriched in ALL inflammatory hallmarks and negative in myogenesis and oxidative phosphorylation. This would suggest that dissecting the noise originating from other stromal cells, effectively revealed additional, fibroblast-specific differences. The results are also in

agreement with the hypothesis of two difference CAF subtypes namely myCAF and iCAF.

To clarify the confusion around the directionality among 166 CAF-markers used earlier, we performed GSEA analysis of newly identified CAF-clusters against these same CAF-markers. Results showed that the markers are indeed preferentially upregulated in human CAFs (Fig3D). This implies that LKB1, NUAK2 and SIK3 actually shows strongest similarity towards human CAFs with respect to these marker genes.

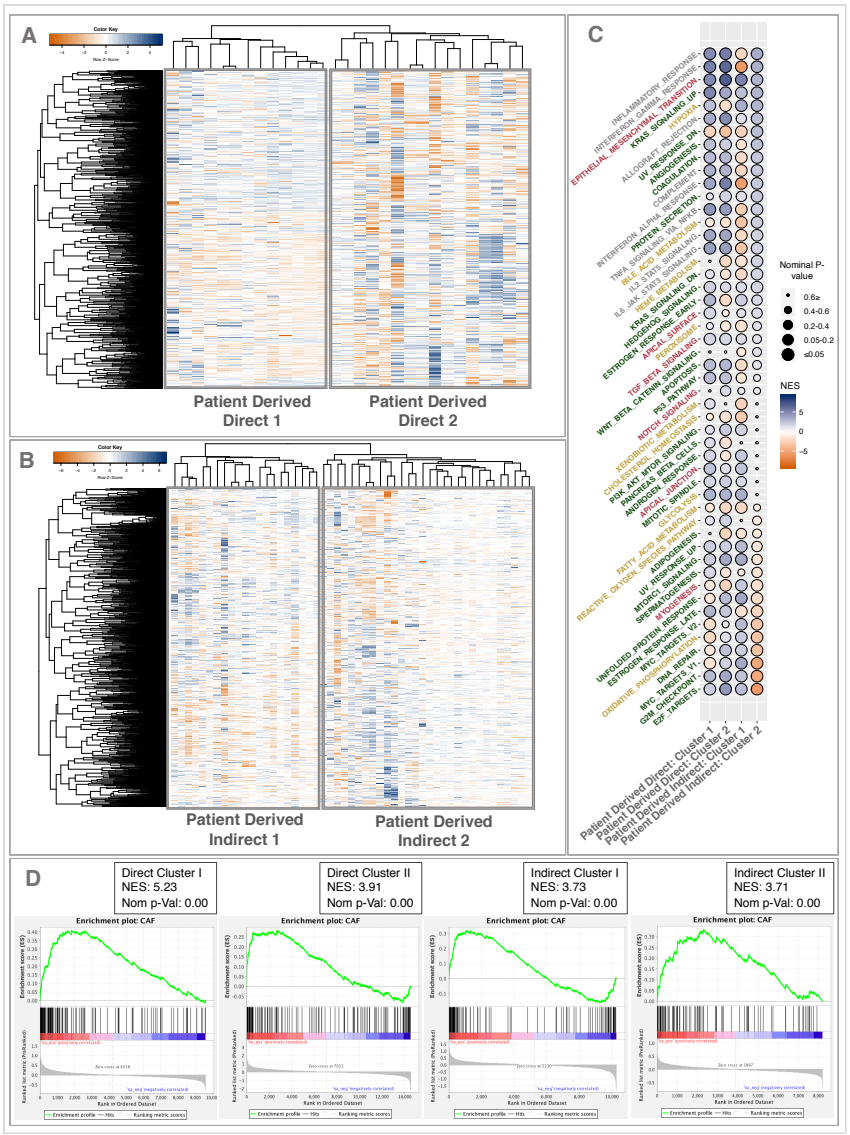
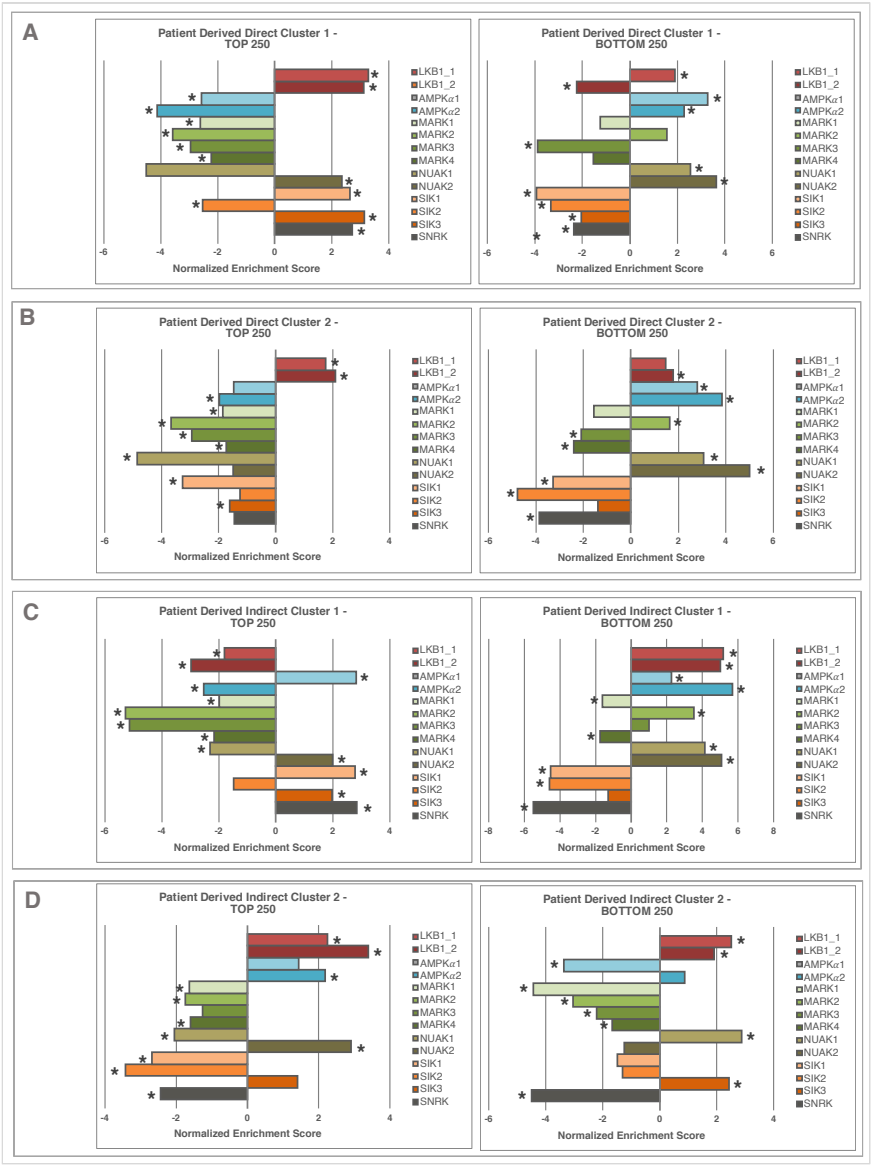


Figure 3. Clustering of Published Stromal Data Sets Dissects CAF-Heterogeneity. A) Unbiased hierarchical clustering of top/bottom 1000 genes of stromal samples directly dissected from patients. B) Unbiased hierarchical clustering of top/bottom 1000 genes of fibroblasts outgrown from patient derived stromal samples. Grey boxes in A) and B) represent the identifies clusters I and II in both. C) GSEA analysis of MSigDB hallmarks of averaged clusters (grey boxes in A) and B)). Color-coding divides hallmarks into inflammatory (grey), metabolic (yellow), myofibroblast (red) and other (green). D) GSEA analysis of patient derived clusters against gene set constructed from curated CAF-markers.

Next we sought to see how the MEF-data compared against the identified clusters themselves. LKB1 datasets represented highest association with direct cluster 1 and indirect cluster 2 (FigS5). This pattern was shared with SIK1, SIK3 and SNRK in the direct, and AMPKα1, AMPKα2 and NUAK2 in the indirect clusters respectively. Importantly, both of the clusters (Direct1 & Indirect2) were low in oxidative phosphorylation and partially glycolytic, which is consistent with initial observations with LKB1-MEFs (Fig2E). In light of this data we conclude that, though not proven a CAF, LKB1-null MEFs preferentially associate together with certain CAF subtypes over others. Namely, LKB1-MEFs are associated with highly inflammatory, low-OXPHOS and low-MYOGENESIS CAF subtypes.



Supplementary Figure 5. GSEA Analysis of LKB1-ARK-MEFs Against Gene Sets Constructed from Patient Derived clusters. A) MEF sequencing samples against Patient Derived **Direct** Cluster 1. B) MEF sequencing samples against Patient Derived **Direct** Cluster 2. C) MEF sequencing samples against Patient Derived **Indirect** Cluster 1. D) MEF sequencing samples against Patient Derived **Indirect** Cluster 2. Asterix signifies statistical significance (nominal p-value ≤ 0.05).

5.4 LKB1 and ARKs Regulate MEF Oxidative Metabolism and Mitochondrial Biogenesis

We have thus far collected two independent transcriptomic evidence of LKB1 functioning in mitochondrial metabolism: unbiased GSEA hallmark analyses and association with low-OXPHOS CAF subtypes. There is also increasing line of evidence supporting LKB1's role as a metabolic caretaker in other cell types [94,95,96,97]. Thus, we hypothesize that LKB1 function as a metabolic regulator is conserved in fibroblasts. More specifically, we asked whether these transcriptomic signatures would translate into physiological responses and further, which one of the kinase substrates might be responsible for it. This metabolic rewiring could potentially help explaining LKB1's role also in ovarian cancer development. There is currently no all-inclusive measure of oxidative phosphorylation or glycolysis. To answer our questions, we began by measuring the MEF oxygen consumption rate (**ORC**) and extracellular acidification rate (**ECAR**) of the knock-down primary cell lines. OCR and ECAR are generally taken as surrogates for oxidative phosphorylation and glycolysis respectively.

Due to the sensitivity of OCR measurement to cell density, and in order to rule out density-dependent effects, we measured it once in low (FigS6B) and twice at high (Fig4A, FigS6A) cell densities. Regardless of plating density, knock-down of LKB1 resulted in reproducibly reduction in OCR levels, reflecting the decreased OXPHOS capacity (Fig4A, upper time course). More specifically, knock-down of LKB1 resulted in decrease in maximum and spare respiratory capacities, having minimal effects in other OCR-measures (Fig4A, Blue barplots). Interestingly the ATP derived through OXPHOS was not altered after LKB1 knock-down, suggesting LKB1-null MEF are still at energetic equilibrium. From the substrates, **AMPKα1** and **MARK3** were the most consistent with LKB1. NUAK1 and SIK1 also showed somewhat similar trends as LKB1 but so far, we only captured 2 replicates from them. AMPKα2 and MARK2 were similar in 2 out of 3 measurements but contrasting in one compared to LKB1. MARK4 and SNRK had increased OCR levels compared to SCR-control.

ECAR levels of LKB1-null MEFs were generally not changed (2/3) but high glycolytic levels were observed in a single measurement (1/3) (Fig4A, lower time course, yellow barplots; FigS6C). Looking for substrates in-between these LKB1-observations we identified **AMPKα1, MARK1** and **SIK-family** as the most similar to

LKB1. Once again, MARK4 and SNRK showed the highest ECAR levels (Fig4A). These results suggested that LKB1 might actually regulate oxidative and anaerobic metabolism through different substrates. Also, MARK4 and SNRK seemed the *strongest* LKB1-independent substrates with respect to metabolic control implicated by both extremely high OCR and ECAR measures. Nevertheless, additional experimentation is required to dissect LKB1-dependent ARK with high confidence.

Reactive oxygen species are known to originate as a by-product of mitochondrial respiration but also by mitochondria-independent means. Regardless of its origin, ROS can potently cause mitochondrial dysfunction and oxidative stress resulting in vicious cycle fueling more and more ROS further reducing the mitochondrial function. Importantly, loss of LKB1 in MEFs is described to induce ROS production [111]. Interestingly, this is independent of AMPKa1/2 [111]. We measured ROS production from knock-down of LKB1 and substrates in MEFs. Importantly, loss-of LKB1 resulted in reproducible ~2-fold increase in ROS (Fig4B, FigS6D). Loss-of AMPKs did not increase the ROS production consistent with earlier report. MARK1, MARK4, NUAK1, SIK2 and SNRK knock-downs increased ROS and thus represents potential candidates for LKB1-antioxidant effectors.

Next, we went to the level of mitochondria to see if loss-of LKB1 or ARKs affects mitochondrial biogenesis. We measured mitochondrial DNA (mtDNA) content and quantified the mitochondrial organelle content. Knocking down LKB1 resulted in 20-40% reduction in cellular mitochondrial DNA content (Fig4C, FigS6E). AMPKα2, MARK2, SIK2 and SIK3 were the most similar in their mtDNA contents. MARK4 and SNRK showed increase in mitochondrial DNA. Somewhat surprisingly, also AMPKα1 showed induction of mtDNA content, which is in agreement with the report describing AMPK as an activator of mitochondrial biogenesis [160].

Similar reduction in mitochondrial protein content (~25%) was also observed in LKB1-MEFs with microscopy (Fig4D). Importantly, we did not observe difference in mitochondrial particle number, suggesting that LKB1-dependent reduction in mitochondrial content is not due to disruption of fission/fusion dynamics. Though not quantified for the ARKs, we did observe considerable fractionation of mitochondria following knock-downs of MARK3 and NUAK1 (FigS7). Also, a different type of

mitochondrial phenotype, a *pearl-necklace* type was observed in AMPKa2-null MEFs (FigS7).

Knock-down efficiencies were evaluated by RT-PCR, at the end of the experiments (FigS8). We didn't observe large differences between the SCR and MOCK-controls in OCR, ECAR, ROS or mtDNA analyses.

Taken together, our data suggests that LKB1 is regulating mitochondrial metabolism and biogenesis in fibroblasts. Moreover, LKB1 may mediate its glycolytic, oxidative phosphorylation and mitochondrial biogenetics effects through distinct substrates. Mitochondrial dysfunction following LKB1-loss was readily observable through multiple levels of inspection. Thus, we feel confident to say LKB1-MEFs are indeed low-OXPHOS. This also clarifies the confusion after seeing conflicting results in LKB1-transcritomes and GSEA (Fig2E).

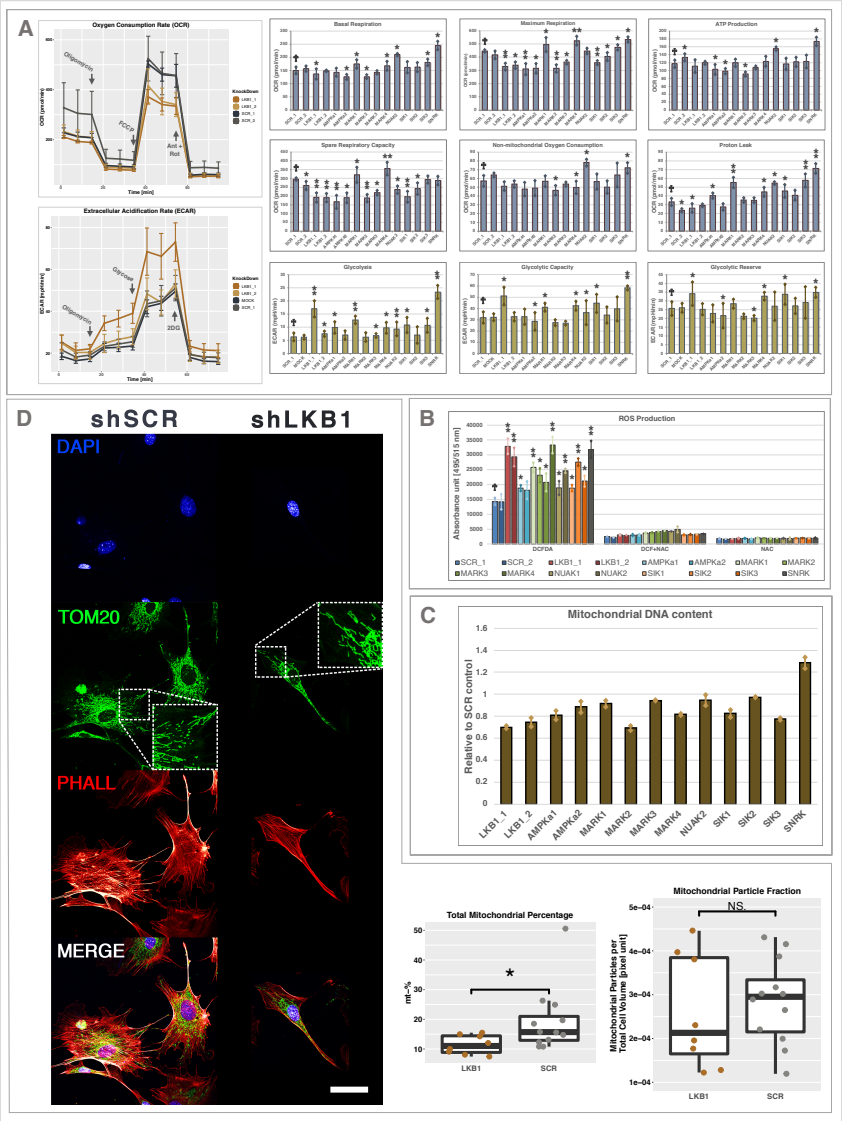
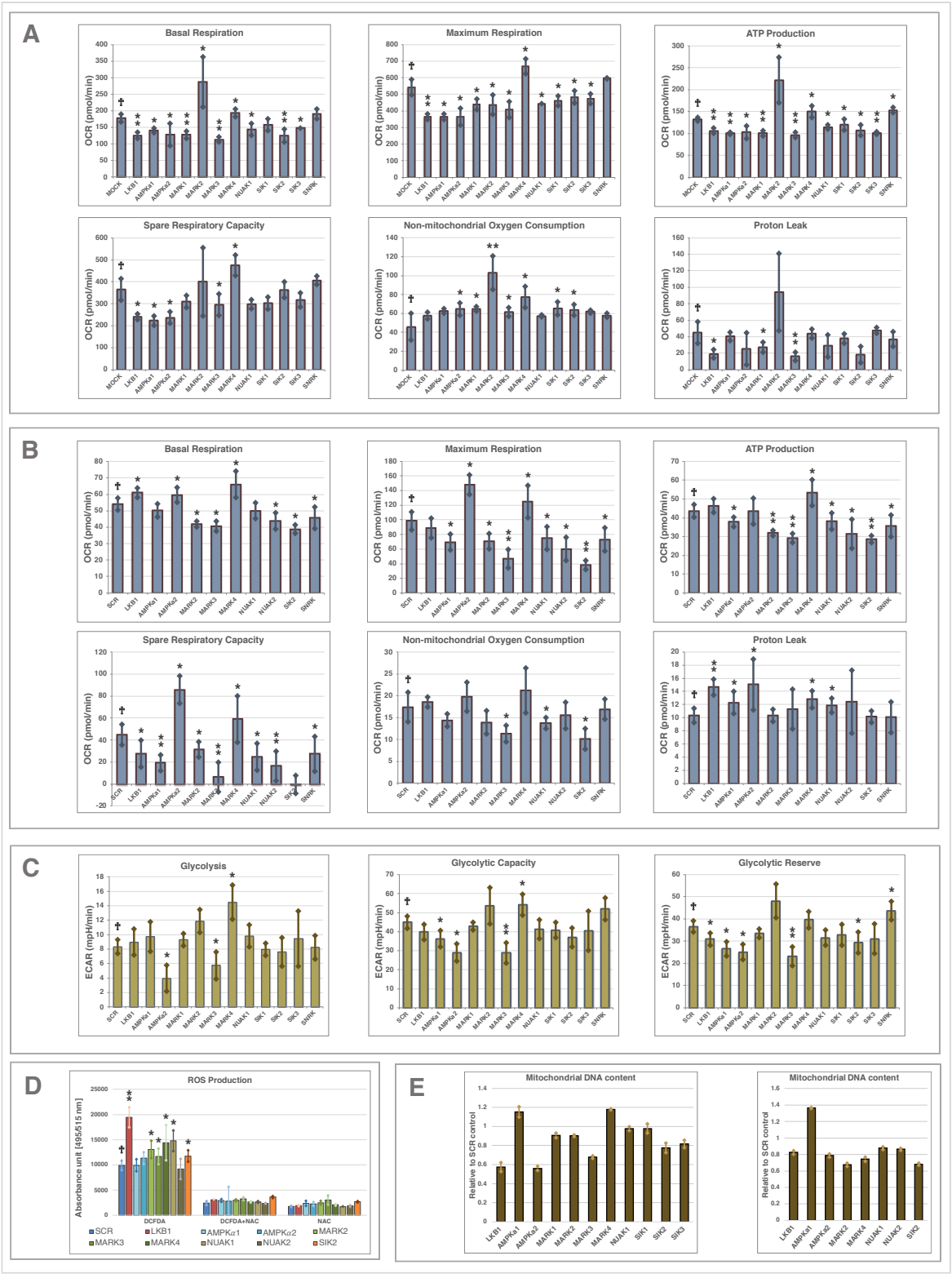
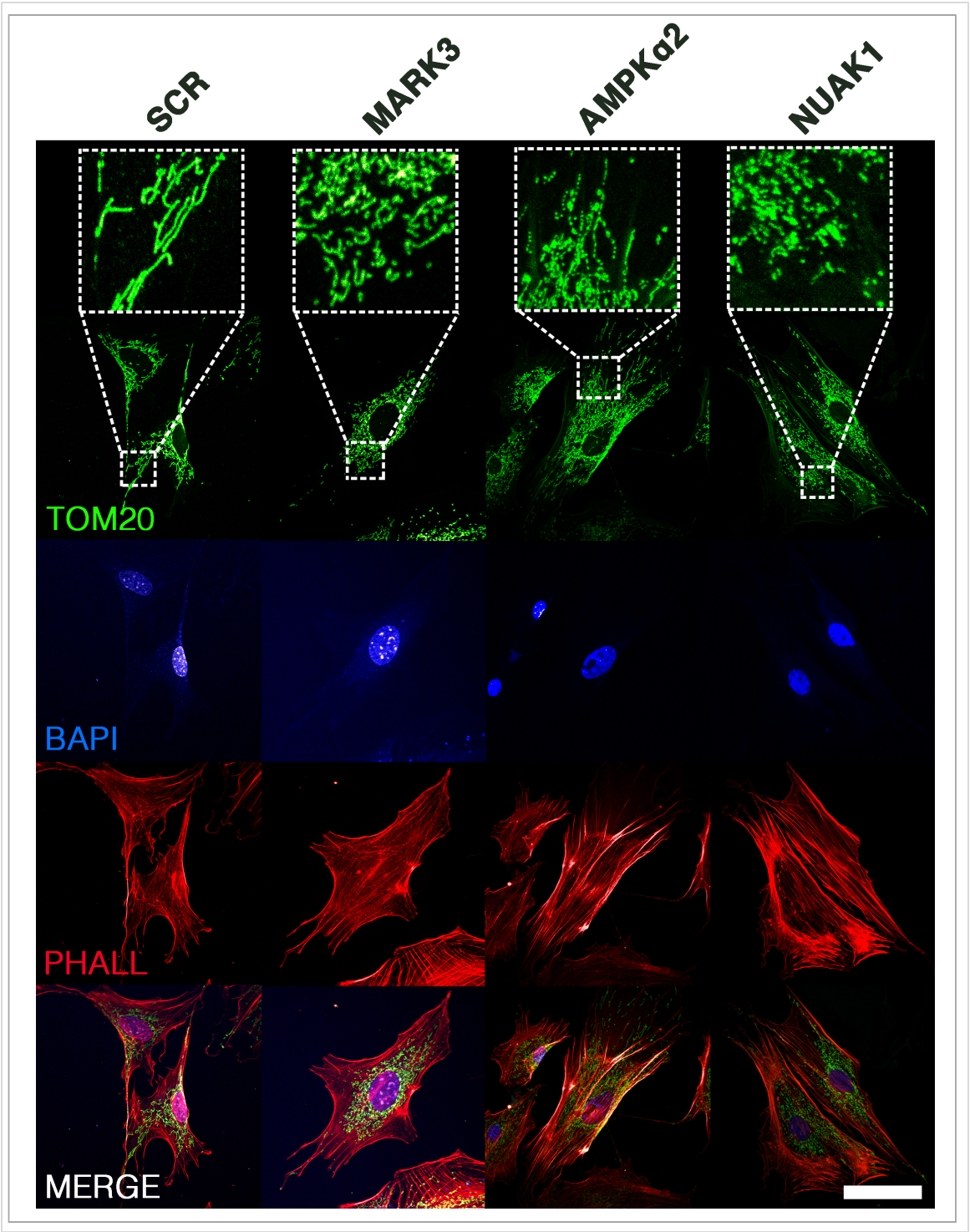


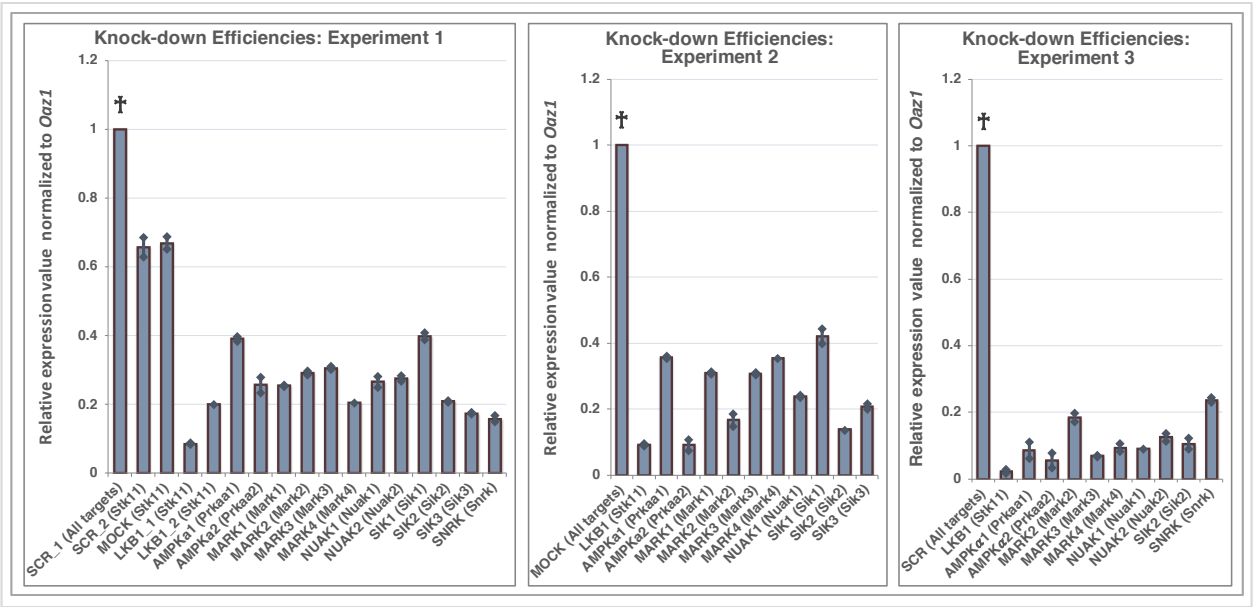
Figure 4. LKB1 and Critical Kinase Substrates Regulate Mitochondrial Metabolism in MEFs. A) Extracellular flux assays of LKB1-ARK-MEFs. Representative time course plots of LKB1 and SCR-/MOCK-controls (2 independent biological replicates from both) in OCR (up) and ECAR (bottom). Barplots representing the OCR (blue) and ECAR (yellow) measures derived from the time course assay. B) Representative barplots from ROS-assay using the same cells as in A). DCF+NAC and NAC represent technical controls. C) Mitochondrial DNA content measurement form the same cells as in A) and B). D) Mitochondrial mass analysis from single pair of LKB1 and SCR cells using the same cells as in A), B) and C). Representative images and quantification of mitochondrial surface area per total area (left) and mitochondrial particle number per total area (right). Each dot in boxplots represents a single cell. Scalebar 50μm. Asterix signifies statistical significance ((*) = p-value ≤ 0.05, (**) = p-value ≤ 0.0001) calculated with unpaired student's T-test. Cross signifies the control samples. No statistics were applied on mtDNA amounts.



Supplementary Figure 6. Data from OCR/ECAR Replicated Experiments. A) OCR measures derived from another **high** cell density flux assay. B) OCR measures derived from **low** cell density flux assay. C) ECAR measures using the same cells as in A). D) ROS-assay using the same cells as in B). Mitochondrial DNA quantification from the same cells as in A)/C) (left) and B) (right) respectively. Asterix signifies statistical significance ((*) = p-value ≤ 0.05, (**) = p-value ≤ 0.0001) calculated with unpaired student's T-test. Cross signifies the control samples. No statistics were applied on mtDNA amounts.



Supplementary Figure 7. Mitochondrial Phenotypes of ARKs. 1st Column representing uniform SCR-control mitochondria. 2nd and 4th Columns highlights the mitochondrial fractionation following knock-downs of MARK3 and NUAK1 respectively. 3rd Column represents the pearl-string phenotype of AMPKα2 knock-down. Cells were the same as in Fig4. Scalebar 50μm.



Supplementary Figure 8. Knock-down Efficiencies of Cells Used in Mitochondrial Assays. Left) Cells used in Fig4. Middle) Cells used in FigS6A, C and E (left). Right) Cells used in FigS6B, D and E (right). Cross signifies the control samples.

5.5 LKB1 Regulates Mitochondrial Metabolism of Ovarian Fibroblasts

Despite *Lkb1*$^{Twist2KO/+}$ was incapable of initiating FRT tumorigenesis, complete stromal deletion of LKB1 succeeds in it [78]. The molecular mechanism is however only partially understood, especially from the perspective of CAFs. We have already provided evidence that loss-of LKB1 or ARKs regulate CAF-associated metabolic changes. To see if similar changes occur in human ovarian fibroblasts, we measured **OCR** and **ECAR** fluxes from LKB1 knock-down iHOF cell lines. As in the case of MEFs, loss-of LKB1 in iHOFs resulted in significant decrease in maximum and spare respiratory capacities and a slight increase in glycolytic responses (Fig5A). Compensating for the reduced OXPHOS, LKB1-null iHOFs increased their glycolysis and glycolytic capacity (Fig5B). Importantly, knocking-down LKB1 from iHOFs didn't increase the proliferation rate compared to SCR control (Data not shown). Thus, this suggests mitochondrial metabolism as a LKB1-mediated mechanism supporting ovarian cancer development.

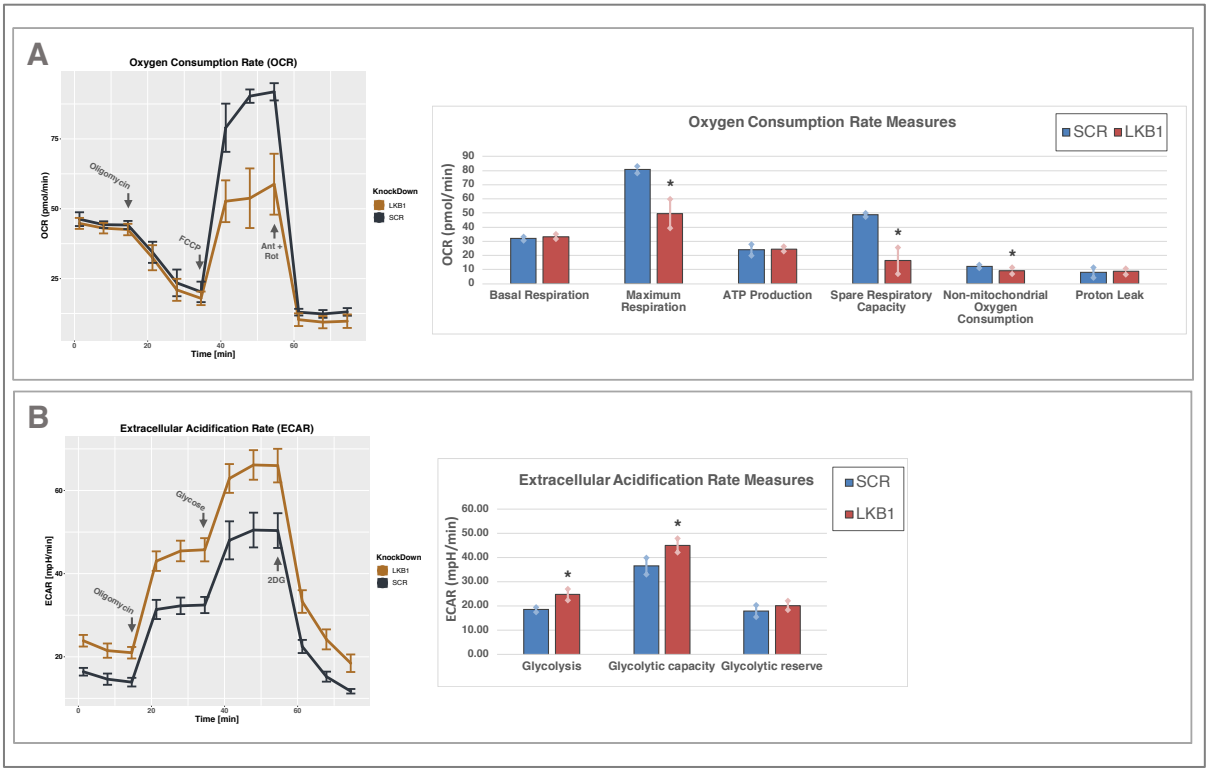
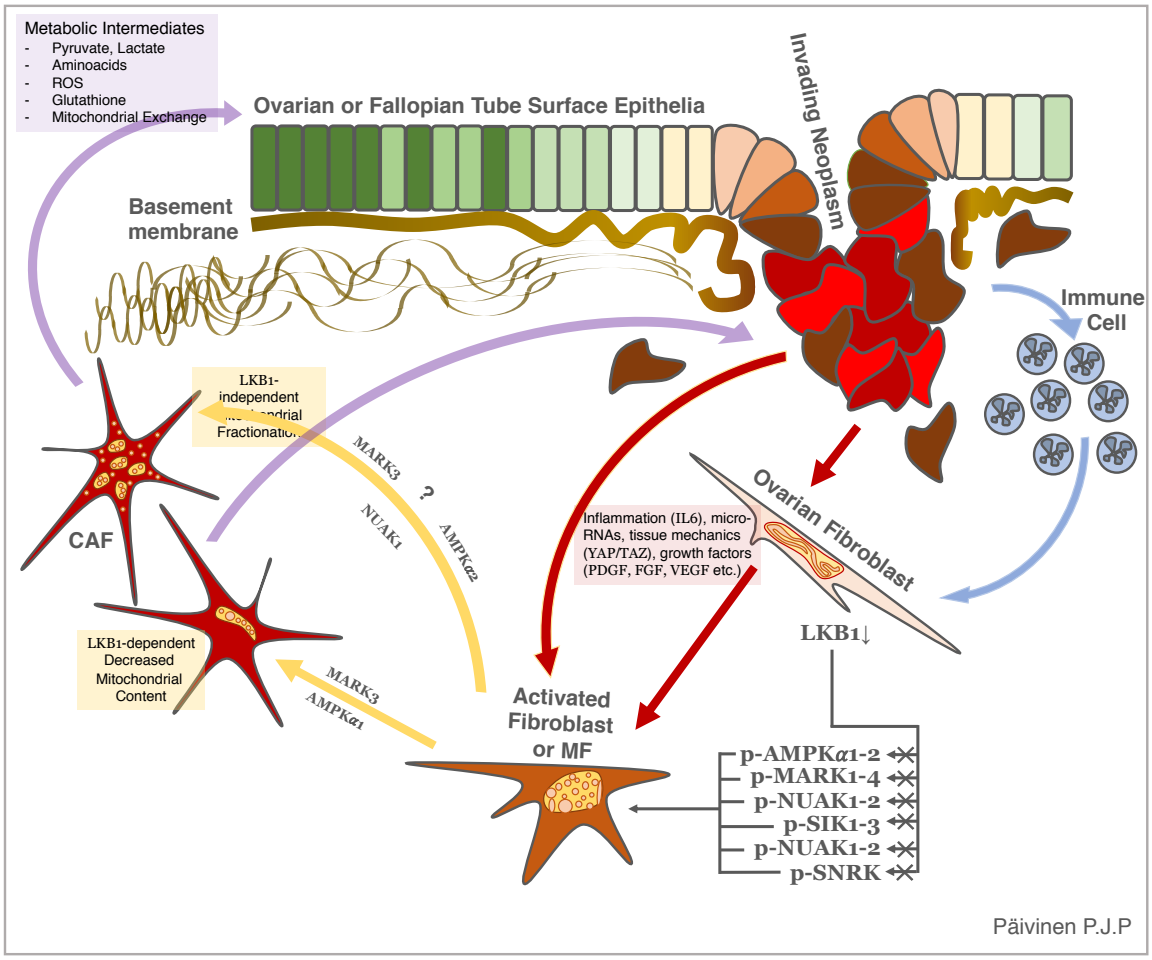


Figure 5. LKB1 Mediates Mitochondrial Metabolism in Human Ovarian Fibroblasts. A) Extracellular OCR assay of LKB1-iHOFs and SCR-control. Representative time course plot (left) and barplots representing OCR measures derived from time course assay (right). B) Extracellular ECAR assay of LKB1-iHOFs and SCR-control. Representative time course plot (left) and barplots representing ECAR measures derived from time course assay (right). Both A) and B) are from same cells. Knock-down was validated in Fig1C (LKB1_3). Asterix (*) signifies statistical significance (p-value ≤ 0.05) calculated with unpaired student's T-test.

Collectively, our data suggests that LKB1 levels are *de novo* down-regulated in human ovarian cancer. Though LKB1-heterozygosity was not sufficient to drive tumorigenesis, near complete loss-of LKB1 in iHOFs resulted in transcriptomic profile reminiscent to ovarian cancer stroma. We also provide evidence that LKB1 regulates cancer-associated metabolic changes also in both murine and human ovarian fibroblasts. This could help explaining the molecular mechanism through which CAFs promote ovarian cancer. These data also suggest more conserved function of LKB1 as a metabolic caretaker in fibroblasts. Whether these metabolic changes are actually causal for cancer initiation or whether they are driven through LKB1 substrates, our data is not strong enough to prove. From the MEF-transcriptomic analyses we can however say that majority of LKB1 functions are driven through small subset of substrates (MARK2, AMPKα2 and NUAK1). Thus, it is plausible to hypothesize (some of) them as actual metabolic effectors of LKB1 that could potentially function also downstream of LKB1 in ovarian cancer.

In light of this data, we propose a model in which **I)** LKB1 function is compromised by yet undescribed mechanism in e.g. ovarian stromal fibroblasts. **II)** This loss-of-function results in deficient activation of LKB1 substrates and **III)** is relayed as mitochondrial dysfunction. **IV)** This again is critical for stromal activation and cancer-associated fibroblast transformation (Model).



Model: Ovarian cancer reduces stromal LKB1 levels. This results in insufficient activation of LKB1's downstream substrates which again causes mitochondrial dysfunction, ROS-production, inflammation, fibroblast activation and transformation into myofibroblast (MF) and/or CAFs. Red arrows indicate direct mechanism by which cancer cells downregulate stromal LKB1. Blue arrows indicate indirect mechanism, in this case through cancer-associated immune cells. Yellow arrows indicate CAF-transformation though both LKB1-dependent and -independent (mitochondrial fractionation). Purple arrows indicate the crosstalk between CAFs and tumor cells through established mechanisms such as aminoacids biosynthesis and lactate shuttle. What is not visible in the model of the signaling from non-invading cancer-cells to the stroma which is also possible. Epithelial color-coding from green to red represents the corresponding increase in malignant potential. Fibroblast intracellular organelles (yellow) represent mitochondria. Brown spiral material underneath the basement membrane represent fibrous ECM which tends to linearize towards the more cancerous region.

6. Discussion

With the data provided, I hope we have convinced the reader that **I)** ARKs contribute to LKB1's functions in hierarchical fashion and that majority of LKB1's functions are covered through its 12 kinase substrates, **II)** Cancer-associated fibroblasts experience profounding heterogeneity which makes their categorization challenging, **III)** LKB1 and some ARK knock-down fibroblasts are activated and share CAF-like phenotypes, **IV)** LKB1 and ARKs regulate stromal oxidative metabolism and lastly **V)** stromal LKB1 regulates functions relevant towards human diseases such as ovarian cancer.

Despite AMPK's historical role as a *pivotal* target of LKB1, the field is gradually beginning to appreciate the spectrum of its other substrates [162]. AMPK may be highly expressed and central in many cellular pathways but as we have already seen, the sheer expression value does not necessarily translate to the functional relevance. This is also why it might be a good idea to include the BRSK1/2 in later transcriptomic-screenings. It would also be interesting to explore the substrates-independent ~30% of LKB1 transcriptome. It is possible that by evaluating the transcriptomes in more physiologically relevant conditions (e.g. energy deprivation) would recover some of this missing fraction. In the lab we have generated transcriptomic data using AMPK-activators. It will be interesting to see how that will compare to LKB1. Additionally, it is possible that some of the remaining 23% is not covered due to paralog-compensatory mechanism. Of note, since LKB1-heterozygosity can already cause neoplasm, it might be worth wile to evaluate the coverage of $LKB1^{-/+}$ MEFs versus complete knock-down.

ARK nomenclature suggests, though hardly proves, inter-substrate functional relatedness. Phylogenetic separation of the substrates is based on protein sequence similarities, which do carry functional information. Functional relatedness can further lead to inter-substrate feed-back/forward loops that can possible affect paralog transcription when another paralog is reduced. We do observe one such loop in case of knocking down AMPKα1, which results boost in α2 expression. This could be taken into account by adding α1/α2 double knock-out to the analysis. Substrate paralogs

also contain additive effects among each other [162]. We have also generated AMPKα1/α2 double knock-out transcriptomic data which is planned to include in the later analyses.

It is interesting to see that AMPKα1 and α2 contribute so much to LKB1's fibroblast signature, even though they cluster far away. Even more, since they are dispensable in GI-polyposis [105]. If it would turn out that AMPKα1/α2 would conduct LKB1's metabolic functions, this would mean the OXPHOS/glycolysis are not critical for polyposis initiation. But this could of course underlie differences between the model systems. Importantly, our model is lacking natural epithelial-to-mesenchyme signaling hub, which plays a role in the living organism. Temporal dynamics in GI-polyposis are relatively long (months/>1 year). Thus, it is more than probable that our brief *in vitro* experimentation is not able to capture real malignant state of polyposis. Rather it measures more natural and homeostatic substrate contribution. It could also be that AMPKα's would mediate other cancer mechanisms aside from initiation. They could for example have role in metastasis and PJS's association to other malignancies.

Outside LKB1, it would be very interesting to look what the substrate's LKB1-independent functions are. Could we e.g. confirm some of the previously identified substrate-specific pathways merely from the transcriptomes? As we have shown, given a substrate, LKB1 maximally covers only 16% of the differentially expressed genes. Naturally, LKB1 sits above the substrate and thus the transcriptomic changes will fade away moving further from the substrate-of-interest. This probably explains why we see such a small LKB1-dependent fraction. Additionally, after seeing this high coverage (77%) in such a heterogenous *in vitro* model, it will be extremely exciting to see what kind of changes in substrate contributions we will see in more clonal/homogenous cell populations such as iHOFs. If a group of substrates could be linked to certain cellular function of LKB1, it would make it easier to rationalize what substrates might be underlying certain phenotype. Here we argue, that underneath one such phenotype, metabolic regulation, there is already a group of substrates emerging. If indeed, AMPKα1 and MARK2 would be the crucial kinases carrying out LKB1's functions, they would serve as plausible candidates for e.g. pharmacological targets. Of course, LKB1's effects could very well be counter balanced through its

independent substrates such as SNRK and MARK4. As is visible from the results, the metabolic behavior is still very noisy, and requires further characterization and rescue-experiments to actually identify the causal substrates.

Despite LKB1-MEFs were exhausted, visible as reductions in maximal and spare respiratory capacities, we do not see any changes in the OXPHOS derived ATP. This might be explained that following LKB1 knock-down, other changes support ATP production such as glycolysis and β-oxidation. Importantly, basal respiration was kept in the level of SCR-control, thus this compensatory ATP production would be anaerobic. Interestingly though, in 2 out of 3 measures we observed less leakiness of protons following LKB1-loss. Thus, we argue that despite loss-of LKB1 and exhaustion of mitochondria, the cell manages to *tighten/restrict* the leakiness of mitochondrial membranes and thus increasing the mitochondrial membrane potential. To prove this hypothesis, easy experiment would be just to measure the Ψ. This would also explain the lack of compensatory glycolysis: since, the cell can handle loss-of LKB1 and following exhaustion by just adjusting its mitochondrial leakiness, it is not necessary to induce glycolysis. Still, since we do observe some degree of increased ECAR in MEFs, this might suggest us an additional backup compensation. To gain insight into this it would be interesting to measure pathways such as β-oxidation/lipid metabolism, glutamine (SEAHORSE Mito Fuel Flex assay) and other aminoacids biosynthetic pathways.

Reactive oxygen species originate in principle through all oxygen related metabolism in the cell. Our data suggests that ROS does not directly correlate with mitochondrial dysfunction. Thus, mitochondria seem to be protected to some extent from intracellular ROS. Data also suggests that mitochondrial dysfunction would be only partially responsible for ROS. Increased ROS could potentially originate through mitochondrial dysfunction or by mitochondria-independent means such as peroxisome and antioxidant dysregulations. In case of MARK4 for example, we observe high OXPHOS and glycolysis, high metabolic activity overall. This could act a potent source of metabolically derived ROS that could in principle relay back to mitochondria and disrupt OXPHOS. Importantly, OXPHOS derived ROS would actually develop inside the mitochondria and has thus shorter distance to the site of

disruption compared to mitochondrial-independent ROS. Loss of LKB1 results in induction of the latter type of ROS. This ROS however conveys to mitochondria, due to extremely high proportions, and only then disrupt the electron transport chain. The fact that loss of LKB1 is still more potent ROS inducer than e.g. MARK2/3 is supporting this idea of two independent but intertwining ROS pathways. To fit this hypothesis to the data, it would mean that probably this excess ROS production would be independent of the 10 substrates. This since none of substrates reach ROS levels of LKB1. This is also suggested by the published report [111]. This would suggest that if LKB1 mediates its OCR effects through AMPKα1, it would be ROS-independent. Importantly, the original report argued that both α1 and α2 were able to rescue the effect, which our data is not strong enough to say. LKB1 is known to activate AMPK-TSC2 mediated cellular damage response following ROS-induction [162]. This could potentially implicate many interesting scenarios where LKB1-AMPK axis would try to provide fail-safe for mTORC1 regulation during e.g. mitochondrial dysfunction. MARK4 and SNRK produced ROS originates probably through LKB1 independent means, suggested by the differences in OCR/ECAR measures. To really dissect the origin of intracellular ROS we could perform mitochondrial specific ROS-assay.

Mitochondrial DNA contains genetic information about the OXPHOS enzymes. Thus, the absolute mtDNA amount sets the limit for mitochondrial components, mass and eventually to function (OXPHOS). In our experiments, we hold on to the model in which mtDNA gets degraded as a result from prior mitochondrial dysfunction. Following mitochondrial impairment, process of coordinated mitochondrial destruction, mitophagy, takes place. Along with the mitochondrial proteins, also the mtDNA gets degraded. This mitochondrial degradation can be observed as e.g. decreased OXPHOS.

Though mtDNA and OXPHOS/glycolysis did not absolutely correlate throughout our experiments, we did observe association of lower mtDNA level and low-OXPHOS in multiple samples. However, the proposed model does not explain how AMPKα1 experience 1.2-1.4 increase mtDNA load, even though the cells are low on OXPHOS. In other words, AMPKα1 could regulate mitochondrial function and

structure independently. It will be interesting to see if the latter is mediated through the reported mechanism [160]. Moreover, since only the functional half is similar to the case of LKB1 suggests that the structural regulation could be mediated through LKB1-independent mechanisms, such as CAMKK2-pathway. Also, seeing this effect only after knocking down α1-subunit, yearns for the effect of double knock-down (DKO). Whether the mtDNA amount would be compensated or potentiated following loss-of additional catalytic subunit could speak in favor or against AMPK-compensatory mechanism. Importantly, in U2OS cells either α1 or α2 reconstitution to DKO-cells was efficient to restore the mitochondrial fission capacity. It is also possible that LKB1 et co. regulated mtDNA through its replication. mtDNA is polymerized by DNA polymerase gamma (*Polg1*). We didn't observe big changes in the *Polg1* mRNA levels in RNA-sequencing though.

We are left puzzled how might MARK4 and SNRK have such a big effect in LKB1-independent cellular metabolism. This was of course suggested in wider perspective at some of the transcriptomic level results. One might speculate that LKB1 promotes oxidative phosphorylation through AMPKs, MARK2/3 and SIK2, whereas MARK4 and SNRK balances this by restraining OXPHOS and glycolysis by LKB1-independent means.

CAF heterogeneity is already established in a variety of human cancers [55,56,64]. Though novel single-cell based approaches will undoubtedly increase the resolution in this field, it is encouraging to see that some of this diversity is possible to capture even by more traditional and easy-access means. As a quick-and-dirty exploration, our CAF datasets seem to capture many relevant biological processes such as myogenesis, metabolism and inflammation. We sought to just take the raw expression values from individual datasets with no further filtering applied. Given a certain gene, adjusting to standard deviation within a cluster may or may not make sense depending on the question. For our purposes we were only interested in the brute ranking of the genes within a cluster so averaging the expression values should work equally fine. If the exact expression value would be important, then normalization would absolutely be in place. Further adjusting of the data might also be possible by means applied from modern single-cell sequencing techniques: by

recording external variables and using them as covariates, could possibly help in reducing the dimensionality of the data. Also including some positive and negative controls in the training such as myofibroblasts, fibroblasts activated by inflammatory cytokines or mitochondrial complex inhibitors, could further help in clustering.

We were surprised how homogenous the direct clusters were in terms of enriched hallmarks. The direct samples were mostly acquired through laser-capture microdissection. Only a handful of FACS sorted fibroblast samples were included. Thus, these clusters represent more or less the total stromal cell heterogeneity: fibroblasts, myofibroblasts, pericytes, immune cells, stellate cells, mesenchymal stem cells etc. This highlights that, though the stromal fibroblasts may differ quite drastically, the stroma as a whole can actually possess striking homology. This even though the that the stroma were dissected form a variety of tissues and tissues specific fibroblast heterogeneity is an acknowledged phenotype in mammalian systems [29]. Most probably this additional layer of stromal heterogeneity could be observed through higher principal components. It might be however possible that regardless of the site of cancer, the stroma that it sculptures would end up looking quite homogenous.

From the indirect clusters it was encouraging to see how to myCAF and iCAF subpopulations were readily visible. In the study describing them, these two subtypes were observed dynamically differentiating between each other, merely through exchanging the culture condition form 2D (myCAF) to 3D (iCAF) [64]. Importantly, in our datasets majority of the indirect samples were cultured in 2D and might thus be expected to represent the myCAF population. Apparently, the differences between the clusters where so fundamental that the enrichment analysis was capable of capturing them.

The observation that LKB1-MEFs actually show high association towards any of the clusters, already argues that the phenotypes that we see there, are associated with human malignancies. Though the data is not strong enough to identify LKB1-MEFs as CAFs, we are positive that they are activated. And by activated, we mean active in CAF-related biological processes. Indeed, the inflammation and energy metabolism are already described as hallmark of loss-of LKB1 in the stroma. Restricting the LKB1 and ARK grouping to curated CAF-markers proposes also

interesting model. It is suggesting that though each member has a prevalence towards certain CAF subtype identified, it is only the LKB1, NUAK2 and SIK3 that show positive correlation with regards the CAF-markers. This would suggest that other ARKs would drive other cellular pathways, outside these CAF-markers. Of course, now since we have generated these new CAF-clusters, it would make sense to update this CAF-marker list into more comprehensive version.

Since all the identified clusters were still CAFs, it was actually surprising to see how few hallmarks were shared between them. EMT, KRAS, HEDGEHOG, TGFβ and protein secretion being the only ones and all of them enriched positively. These could be thus utilized as a more general characteristic of CAFs. Instead of using a single molecular marker, such as αSMA, one should rather probe at least a single marker from all the above-mentioned hallmarks. It would also make sense trying to classify the genes responsible for the divergence of the clades. By k-means clustering it might be possible to virtually dissect the biologically active pathways in separate clusters. This would also help in identifying single molecular markers for different biological processes such as the above-mentioned hallmarks.

We did not include data on other, non-cancer-but-activated fibroblasts we collected. Nevertheless, in the future they might help dissecting similar associations as in case of CAFs. Also, with the help of more controlled *in vitro* experiment it might be possible to investigate the precise molecular mechanism in detain. They could also be used in attempt to dissect the biological process activated in specific CAF-dataset. For example, one could generate an IL-6 gene set form *in vitro* conditioning of fibroblasts with IL-6. This can be probed against all the CAF-datasets to see which of them seem high or low in IL-6 responsive genes. This way we hope we can further develop our collection of data into more versatile tools for people to use to quickly see how their data compares against the published.

There have been numerous attempts figuring out LKB1's role in reproductive tract development due to increased incidence of ovarian, endometrial and cervical in PJS patients. Still, there is not a single report explaining LKB1's role in ovarian cancer specifically. We have shown for the first time that endogenous level of LKB1 is decreased in the stroma of human HGSOC and it is unchanged in the tumor epithelia.

We also provide evidence suggesting loss-of LKB1 in ovarian fibroblasts results in CAF-like metabolic phenotype.

The fact that we did not observe malignant transformation in *Lkb1*$^{Twist2KO/+}$ mice ovaries, fallopian tubes of uterus is in agreement with earlier reports. It is likely that rather functioning as a driver of initiation, stromal loss of LKB1 enhances the cancer cell survival. It is also possible that the low sample number was not able to capture the phenotype. Indeed, whole-body heterozygous mice are reported to exhibit endometrial cancer only in ~30% of cases [79]. Since LKB1 heterozygosity was not fully penetrant in cancer development, further speaks for additional driver mutations. Complete loss of LKB1 in the endometrium is efficient to induce fully penetrant endometrial adenocarcinoma [79]. This however does not happen in human patients.

Looking closely at the levels of changes in different experiments gives us also important perspective. Looking at the stroma as a whole, we observe ~50% reduction at mRNA level. At the level of fibroblasts this is about 40% (only a single sample). Since we have no information about the stromal cell composition of ovarian cancer, we are not able to say if LKB1 is reduced in other cells also. Still, since ovaries are naturally fibroblast-rich organ it is plausible that this holds true also in case of cancer. Thus, this slight increase in LKB1 levels could originate from the other stromal cells in which LKB1 level in retained. Thus, the level where fibroblast-specific LKB1wil set in ovarian cancer remains elusive.

In our iHOFs the knock-down efficiency reaches about 20-40%, which hits slightly harder compared to actual CAFs (above). Thus, it is important to repeat the experiments with either *heterozygous* iHOFs or with patient cell lines. This could also provide important evidence not only for the case of HGSOC but also PJS: to date there has been no direct (*in vivo*) or indirect (*in vitro*) evidence of mitochondrial dysfunction of fibroblasts in PJS-models. Speculation around metabolic changes in PJS fibroblasts are based on association analyses (*in silico*). Encouraging data is reported from other stromal cells though (hematopoietic and T-cell).

Nevertheless, after seeing such a good coverage between the LKB1-responsive genes and OVCA-CAFs it is difficult to question the association between them. Interestingly though, many of the ECM components were observed downregulated in the LKB1-iHOFs whereas they were upregulated in the OVCA

stroma. This might be due to number of reasons. One explanation for this might come from the culturing conditions: LKB1-null fibroblasts secrete different ECM components depending on whether plated on plastic or situating in their native environment. A more mechanistic idea comes from LKB1's function as a TGFB inhibitor. Reduced TGFB-signaling would be observed as decrease in ECM component expression. We didn't however see big changes in TGFB-signaling component (SMAD9, TGFBI, TGFBR3) so further investigation on this is in place.

Another interesting finding regarding endometrial loss-of LKB1 was that it induced stromal expansion of αSMA expressing myofibroblasts (CAFs). Interestingly, we did not observe such increase in proliferation in LKB1-null iHOFs. This highlights the role of tumor microenvironment. Induction hyperproliferative program would require other cells types and their signaling components. Still, metabolic changes were readily observable after loss-of LKB1 alone. This suggests LKB1's role as a metabolic caretaker, in fibroblasts as-a-whole, seems highly conserved.

Whether *de novo* LKB1 downregulation is achieved though paracrine regulation of genetic silencing is not known. OVCA stroma is reported not to be mutated [51]. This further supports the idea that some of the tumor microenvironmental phenotypes might decay once the epithelial effectors are removed. Thus, also metabolic rewiring could be potentially reverse merely by identifying the critical signaling component. Of course, stroma might face epigenetic pressure, and through that, change its gene expression permanently even without mutational load. Furthermore, fibroblasts originating through EMT might experience mutations, and thus, selective advantage and expand.

A question that is still completely open is through which of the substrates LKB1 might conduct its effects in HGSOC. Many of ARKs have been discovered to play a role in ovarian cancer development but mostly from the epithelial point-of-view and the LKB1-dependency is hardly proven.

7. Future Directions

We are planning to explore the AMPKα1/2 double knock-out MEF dataset against current data. With it we hope to dissect some interdependencies and feed-back/-forward mechanisms. Luckily these data also include energy stress samples that can help us to understand how the mitochondrial mechanisms are behaving in more physiologically active circumstances. We will consider sequencing of BRSKs as the shear expression levels in MEFs does not seen to justify ruling them out. We will also plan rescue experiments in LKB1-deficient MEFs with add-back of at least AMPKα1, MARK1/3 and SIKs. We will also try mitochondrial experiments with LKB1-null MEFs with treatment of AMPK-activators (AICAR, A76, metformin etc.). We are also planning rescue experiments with chemical agonists/antagonists of other substrates. We would also want to look into the mitochondrial fission and fusion dynamics in addition to brute mass. There are currently plenty of good antibodies suitable for this task. Also, we are considering performing the mitochondrial assay with heterozygous deletion of LKB1: currently we have primary intestinal fibroblasts and MEFs suitable for the task. We are also considering similar type of assay in iHOFs or primary HOFs. We are also considering looking more into detail how are pathways such as β-oxidation, nucleotide metabolism and gluconeogenesis within MEF-system. We will aim to further filter, fitting and merging the publicly available datasets with the aim to improve the handling of large deviation between among the samples. With the expertise from single-cell technologies we will try to normalize the samples according to different covariates. Currently available cancer transcriptomic databases are primarily only epithelial samples. Thus, we find this type of cancer stroma repository particularly interesting.

8. Methods

8.1 Cell Lines and Cell Culture

Mouse embryonic fibroblasts were generated by dissecting E13.5 wild type embryos on petri dish. Tissue pieces were trypsinized into single-cell suspension at +37°C. Media was changed shortly after trypsin neutralization, thus flushing unattached epithelial cells away. MEFs were grown in DMEM (10% FBS, 5% L-glutamine, 5% Pen/Strep = $DMEM_{comp}$) to confluency and stored in liquid nitrogen. **Human ovarian fibroblasts** (ScienCell™, #7330) were grown on fully supplemented fibroblast media (ScienCell™, #2301) on uncoated plates. Immortalization was achieved through transfection of C-terminal tp53. **HEK293FT** cells were used for virus production. 293s were maintained in $DMEM_{comp}$. **Patient cell lines** were grown in fibroblast media without immortalization. Cells were culture in standard conditions (+37C°C).

8.2 Lentivirus Production and Transfection

Virus production: HEK293FT cells were transfected using Lipofectamine2000® (Thermo Fisher Scientific, 11668019) following manufacturer's transfection protocol for 6/10cm culture plates. For lentiviral constructs and viral particle sequences, refer to **Appendix II**. pLKO.1 expression vector, pCMV-dR8.9 and VSVG were mixed according the RNAi Consortium recommendations (http://www.broadinstitute.org/rnai/public/resources/protocols). Transfected cells were grown 48 hours before virus containing media was collected. Media was filtered through 20μm filter. Viral vials were stored in -70°C.

Transduction: $2x10^5$ MEFs/HOFs were plated in 6-wells 24 hours prior transduction in 2ml of $DMEM_{comp}$/fibroblast media. 24 hours post plating 1ml of media was removed and supplemented with viral media along with 8μg/ml of polybrene (Merck Millipore, TR-1003-G). 24 hours later all media was changed to fresh. 24 hours later (48 hours post plating) cells were split 1:2 and adding 1μg/ml of puromycin for selection. 48 hours later cells were split once more and puromycin was added at higher, 2μg/ml concentration. For prolonged culture cells were maintained in 2μg/ml puromycin.

8.3 RT-PCR RNA was extracted using Macherey-Nagel NucleoSpin® RNA Plus extraction kit (740984.250) following manufacturer's protocol. RNA concentration was measured with Nanodrop ND-1000 spectrophotometer (Thermo Fisher Scientific, ND-1000). 400ng of RNA was reverse transcribed into cDNA using Taqman reverse transcription kit (Applied Biosciences). 20ng of cDNA was applied for each RT-PCR reaction master mix (KAPA, SYBR® FAST, KK4617). Measurement was performed with StepOnePlus Real-Time PCR system (Applied Biosystems). No-template controls were included for each of the probes and no-reverse-transcriptase controls in case the RNA concentration was seemingly low (<20ng/μl). For PCR primer sequences, refer to **Appendix I**.

8.4 Western-Blotting Protein samples were collected by lysing the cells in SDS-lysis buffer (250mM Tris-HCl, 2.5V/V-% SDS, 0.5 mM PMSF, 10mM Glycerol-2-phosphate, 5mM NaF, 1mM DTT, 1 μg/ml Leupeptin, heated +98°C). Samples were incubated 10 minutes in +98°C and passed through 25G needle 10 times. Samples were centrifuged 30 minutes at full speed and the supernatants were transferred into clean tubes. Protein concentration were measured with DC™ protein assay (BioRad) FLUOstar Omega spectrophotometer (ORDIOR). Samples were denatured in SDS-loading buffer with bromophenol blue, and heated with another 5 minutes in +98°C.

Proteins were separated on SDS-page gel and transferred to nitrocellulose membrane. Protein loading was assessed with Ponceau staining. Membranes were blocking in skim milk for 1 hour (5%-w/V in 0.05%TBS-Tween-20® = milk). After quick was in TBST membranes were incubated in primary antibodies over-night at +4°C (AMPKα1, Abcam, ab110036; AMPKα2, Abcam, ab3760; Lkb1, Abcam, ab15095; phospho-ACC, Cell Signaling, #3661L; αSMA, Abcam, ab5694; CAV-1, Cell Signaling, #D46G3; MMP-9, Cell Signaling, #13667; NDUFB8, Abcam, ab110242; Vinculin, Cell Signaling, #4650). After washing with TBST, membranes were incubated in HRP secondary antibodies diluted in milk for 1 hour in room temperature. After washing, membranes were incubated in chemiluminescent solution (ECL) for 3 minutes following luminescent detection in dark room with autoradiograph films.

8.5 RNA-sequencing and GSEA Analysis RNA quality was assessed with Bioanalyzer 2100 (Agilent). Only samples with RIN-value above 10 were accepted for library preparation. Library preparation (NEBNext® Ultra™ Directional RNA Library Prep Kit for Illumina®), sequencing (Illumina Next Seq500) and data demultiplexing, were done by FuGU (Biomedicum Functional Genomics Unit). Messenger RNA was enriched using poly-A beads. All together 1μg of RNA was sequenced per sample.

FASTQ files were processed through standardized RNA-sequencing pipeline (**Appendix IV**). Raw reads were mapped to mouse reference genome (GRCm38.p6) with STAR [163]. Mapped reads were counted using htseq [164]. Read counts were imported to R environment where differential expression analysis was conducted using limma and edger software's (**Appendix V**) [165,166]. Principal component analysis was performed with limma and hierarchical clustering with heatmap2. Unix pipeline was conducted in CSC's (IT-Center for Science, Finland) TAITO environment.

For GSEA ((V.4.0.3)) [167], human orthologs were fetched with Biomart [168]. Ranked gene lists were compared against hallmark gene sets (V.7) from Molecular Signatures Database (MSigDB).

8.6 Published Datasets Published stromal datasets were fetched from Gene Expression Omnibus (NIH-NCBI) and ArrayExpress (EMBL-EBI) repositories. **For microarray data**, if functional annotation data was available, Affymetrix® (Thermo Fisher Scientific) raw CEL-files were uploaded into R-environment. Raw data was background corrected, RMA-normalized, log-transformed, annotated and analyzed according to **Appendix VIII** (Applied from http://homer.ucsd.edu/homer/basicTutorial/affymetrix.html). For other arrays and if no annotation files were found, differential expression was conducted with GEO2R browser analyzer inside GEO. Results were uploaded into R. Microarray probes were filtered according to the adjusted p-value. Gene symbols and orthologs were fetched using biomaRt (Bioconductor, **Appendix IX**). **For sequencing data**, if raw dead counts were not supplied to the repository, they were fetched using SRA-toolkit (**Appendix IV, V**). Sequencing pipeline was the same as above: STAR mapping to

human (GRCh38.p12)/mouse (GRCm38.p6) reference genomes, read counting with htseq and differential expression in R with limma/edger.

For clustering analyses, single datasets were merged into combined data (direct and indirect separately, **Appendix X**). 1000 Highest upregulated and 1000 lowest downregulated genes were included into the clustering analyses which was conducted with heatmap2 in R. After cluster identification, subcluster datasets were re-merged and the expression values were averaged across into a single representative gene set (**Appendix XI**).

8.7 Extracellular Flux Assay 24 hours prior measurement cells were plated on Seahorse® 96-well plates (Agilent, Seahorse XFe96 FluxPak, 102416) in normal growth media. For high cell density OCR and ECAR measurement $3x10^4$ and for low cell density OCR $1x10^4$ cell were used. At the day of measurement, cells were washed with PBS and supplemented with 80μl of supplemented Seahorse® media (pH7.4). For ECAR, glycose was omitted from Seahorse® media and the cells were starved for 1 hour. OCR and ECAR cells were incubated in +37°C, CO_2-free incubator (VWR, INCU-Line) for at least 1 hour before measurement. Injections were calibrated in XF calibrant solution (Agilent, 100840-000). Measurement protocol for OCR: 3x baseline, 3x oligomycin (O4876), 3x FCCP (C2920) and 3x rotenone (557368) and antimycin (A8674). Measurement protocol for ECAR: 3x baseline, 3x glucose (A2494001), 3x oligomycin and 3x 2DG (D6134).

After measurement, wells were stained with crystal violet for 30 minutes and washed equally 5 times with 100 μl of PBS. Stained plates were scanned, and crystal violet stain was quantified using ImageJ (v1.51s). Seahorse results were normalized according to the crystal violet. Time course figures were generated with R, barplots with Microsoft Excel. Knock-down efficiencies were obtained from parallelly grown cells at the end of the experiment.

8.8 ROS-Assay 24 hours prior measurement, $1x10^4$ cells were plated on 96-well plates (PerkinElmer, ViewPlate-96) in normal growth media. At the day of measurement, wells were divided into treatments: DCFDA (2',7'-dichlorodihyfrofluorscein diacetate), DCFDA+NAC (N-acetyl cysteine), NAC and 'no-

treatment'. NAC-containing wells were treated with 10 mM NAC in PBS for 2 hours. NAC was removed and cells were washed once with PBS. DCFDA-containing wells were treated with 1μM DCFDA in PBS for 20 minutes. All wells were finally washed with PBS and incubated for 60 minutes in +37°C. Measurement was performed with EnSpire spectrophotometer (PerkinElmer) using 495/515nm filter. Blank corrected values were plotted using Microsoft Excel. DCFDA+NAC and NAC results functioned as additional controls.

8.9 mtDNA DNA was extracted using QIAGEN Blood & Tissue Kit (QIAGEN, 69506). DNA concentrations were measured using Nanodrop. Mitochondrial DNA content was measured using 12S ribosomal RNA (12SrRNA) as a mtDNA marker and RBM15 as an extra-mitochondrial control. Primer and DNA were mixed and measured using above RT-PCR protocol. Fraction of mtDNA versus overall DNA was calculated as suggested in [169]. For specific primer sequences, refer to **Appendix II**.

8.10 Immunohistochemistry **For immunofluorescence of MEFs**, cells were washed with PBS, fixed with 4% PFA in PBS for 30 minutes and washed with PBS for 4 times. Cells were permeabilized with 0.3% Triton X-100 in PBS for 10 minutes. Cells were blocked with blocking buffer (10%-goat serum in 0.3% Triton X-100 in PBS) for 1 hour. Blocking buffer was removed and replaced with primary antibody (Tom20, Santa Cruz, sc-11415; Phalloidin-Alexa Fluor™-594 conjugated, Thermo Fisher Scientific, A12381) diluted in blocking buffer overnight in +4°C. Cells were washed 3 times with PBS and incubated in secondary antibodies (Thermo Fisher Scientific, Alexa Fluor™ Goat-α-Rabbit 488) diluted in blocking buffer. Cells were washed 3 times with PBS and incubated in Hoechst (1mg/ml) for 5 minutes. Cells were washed 2 times with PBS and mounted using Mowiol. **For immunohistochemical staining of paraffin embedded mouse reproductive tracts**, mouse ovaries were obtained from CO_2 euthanized mice. Excess fat was removed under stereomicroscope and FRTs were incubated in 4% PFA-PBS in +4°C, overnight. Next day, FRTs were placed in 70% ethanol or washed with PBS and placed into paraffin processing cassettes. Paraffin blocks were serially sectioned

using microtome (MICROM, HM 355 S). 5μm thick sections dried overnight in 37°C platform. Prior deparaffinization series, paraffin was melted by briefly incubating sections on +60°C platform. Antigen retrieval was performed by incubating sections in citrate buffer pH6 or Tris-EDTA pH9 (Dako, S169984-2 and S236784-2) in 95°C for 20 minutes following 30 minutes of cooldown. Sections were quickly washed in dH_2O. Endogenous peroxidases were inactivated by incubating sections in 3%-H_2O_2 diluted in PBS for 10 minutes following 2 washes in dH_2O. Sections were blocked by incubating in TNB blocking buffer (TSA® Blocking Reagent diluted in 0.1%-TBS-Tween20, PerkinElmer) for 1 hour in ambient temperature. Blocking buffer was replaces with primary antibody diluted (Lkb1, CST, #13031; Vimentin, DAKO, M0725; αSMA, Abcam, ab5694; phpsho-SMAD3, Abcam, ab51903) in blocking buffer and incubated in +4°C, overnight. Sections were washed 3 times in washing buffer (0.1%-Tween20 in TBS). **For IHC,** 3 drops SignalStain® Boost Reagent (Cell Signaling Technology, #8114) was applied and incubated for 30 minutes in ambient temperature. Sections were washed 3 times in washing buffer. Detection was done using DAB substrate kit (Cell Signaling Technology, #8059). Sections were counterstained using haematoxylin, dehydrated and mounted. **For immunofluorescent staining of paraffin sections,** endogenous peroxidase depletion step was omitted. Primary antibodies (TdTomato, SCIGENE, AB8181-200; E-Cadherin, SCT, #3195) were recognized using Alexa Fluor conjugated secondary antibodies (Donkey anti-Goat, Alexa Fluor™-488/-594) and mounted section were mounted using ProLong™ Diamond (Thermo Fisher Scientific) mountant.

8.11 Microscopy MEF mitochondrial content was imaged using Leica TCS SP8 CARS confocal microscope (Biomedicum Imaging Unit). Paraffin sections scanned using 3DHISTECH Pannoramic 250 FLASH II digital slide scanner (Genome Biology Unit, HiLIFE, Faculty of Medicine, University of Helsinki, and Biocenter Finland). Scanned slides were analyzed using HistoQuant software (3DHISTECH). Images were exported and panels were generated using Adobe Photoshop.

8.12 Patient Cells and Mouse Model Human samples were a gift from Pietilä E. (Genome-Scale Biology, Research Programs Unit, University of Helsinki, Helsinki,

Finland). Mouse samples were supplied by Doménech-Moreno E. (HiLIFE-Helsinki Institute of Life Science, University of Helsinki, Helsinki, Finland) and were collected in conjunction with polyposis analyses.

9. Acknowledgments

I wish to thank S. Tripathi and T.P. Mäkelä for scientific and experimental supervision throughout the course of the study. I thank E. Pietilä and E. Domènech-Moreno for supplying experimental material. I also want to thank all of my other mentors, Yang Y., Yan Y. and Gao Y. for giving me such a rich working environment, which has truly shaped me as the scientist I am now.

I give special thanks my family and friends for supporting me through in my studies in Helsinki and else-where. I also thank you Elli for all the possible support you have given me.

10. References

1. Skalli O, Ropraz P, Trzeciak A, Benzonana G, Gillessen D, Gabbiani G. (1986), A monoclonal antibody against alpha-smooth muscle actin: a new probe for smooth muscle differentiation. *J Cell Biol*, **103(6)**, 2787-2796.
2. Lorenz K, Sicker M, Schmelzer E, Rupf T, Salvetter J, Schulz-Siegmund M, Bader A. Multilineage differentiation potential of human dermal skin-derived fibroblasts. *Exp Dermatol*. 2008 Nov;17(11):925-32.
3. Ubil E, Duan J, Pillai IC, Rosa-Garrido M, Wu Y, Bargiacchi F, Lu Y, Stanbouly S, Huang J, Rojas M, Vondriska TM, Stefani E, Deb A. Mesenchymal-endothelial transition contributes to cardiac neovascularization. *Nature*. 2014 Oct 30;514(7524):585-90.
4. Denu RA, Nemcek S, Bloom DD, Goodrich AD, Kim J, Mosher DF, Hematti P. Fibroblasts and Mesenchymal Stromal/Stem Cells Are Phenotypically Indistinguishable. *Acta Haematol*. 2016;136(2):85-97.
5. Soundararajan M, Kannan S. Fibroblasts and mesenchymal stem cells: Two sides of the same coin? *J Cell Physiol*. 2018 Dec;233(12):9099-9109.
6. Anttila MA, Tammi RH, Tammi MI, Syrjänen KJ, Saarikoski SV, Kosma VM. High levels of stromal hyaluronan predict poor disease outcome in epithelial ovarian cancer. *Cancer Res*. 2000 Jan 1;60(1):150-5.
7. Costa-Almeida R, Soares R, Granja PL. Fibroblasts as maestros orchestrating tissue regeneration. *J Tissue Eng Regen Med*. 2018 Jan;12(1):240-251.
8. García-Palmero I, Torres S, Bartolomé RA, Peláez-García A, Larriba MJ, Lopez-Lucendo M, Peña C, Escudero-Paniagua B, Muñoz A, Casal JI. Twist1-induced activation of human fibroblasts promotes matrix stiffness by upregulating palladin and collagen α1(VI). *Oncogene*. 2016 Oct 6;35(40):5224-5236.
9. Chaudhuri O, Koshy ST, Branco da Cunha C, Shin JW, Verbeke CS, Allison KH, Mooney DJ. Extracellular matrix stiffness and composition jointly regulate the induction of malignant phenotypes in mammary epithelium. *Nat Mater*. 2014 Oct;13(10):970-8.
10. Calvo F, Ege N, Grande-Garcia A, Hooper S, Jenkins RP, Chaudhry SI, Harrington K, Williamson P, Moeendarbary E, Charras G, Sahai E. Mechanotransduction and YAP-dependent matrix remodelling is required for the generation and maintenance of cancer-associated fibroblasts. *Nat Cell Biol*. 2013 Jun;15(6):637-46.
11. Hinz B, Mastrangelo D, Iselin CE, Chaponnier C, Gabbiani G. Mechanical tension controls granulation tissue contractile activity and myofibroblast differentiation. *Am J Pathol*. 2001 Sep;159(3):1009-20.
12. Desmoulière A, Geinoz A, Gabbiani F, Gabbiani G. Transforming growth factor-beta 1 induces alpha-smooth muscle actin expression in granulation tissue myofibroblasts and in quiescent and growing cultured fibroblasts. *J Cell Biol*. 1993 Jul;122(1):103-11.
13. Tomasek JJ, Gabbiani G, Hinz B, Chaponnier C, Brown RA. Myofibroblasts and mechano-regulation of connective tissue remodelling. *Nat Rev Mol Cell Biol*. 2002 May;3(5):349-63.
14. Di Guglielmo GM, Le Roy C, Goodfellow AF, Wrana JL. Distinct endocytic pathways regulate TGF-beta receptor signalling and turnover. Nat Cell Biol. 2003 May;5(5):410-21. Erratum in: *Nat Cell Biol*. 2003 Jul;5(7):680.

15. Shi Y, Massagué J. Mechanisms of TGF-beta signaling from cell membrane to the nucleus. *Cell*. 2003 Jun 13;113(6):685-700.
16. Vincent T, Neve EP, Johnson JR, Kukalev A, Rojo F, Albanell J, Pietras K, Virtanen I, Philipson L, Leopold PL, Crystal RG, de Herreros AG, Moustakas A, Pettersson RF, Fuxe J. A SNAIL1-SMAD3/4 transcriptional repressor complex promotes TGF-beta mediated epithelial-mesenchymal transition. *Nat Cell Biol*. 2009 Aug;11(8):943-50.
17. Kim Y, Kugler MC, Wei Y, Kim KK, Li X, Brumwell AN, Chapman HA. Integrin alpha3beta1-dependent beta-catenin phosphorylation links epithelial Smad signaling to cell contacts. *J Cell Biol*. 2009 Jan 26;184(2):309-22.
18. Koenig A, Mueller C, Hasel C, Adler G, Menke A. Collagen type I induces disruption of E-cadherin-mediated cell-cell contacts and promotes proliferation of pancreatic carcinoma cells. *Cancer Res*. 2006 May 1;66(9):4662-71.
19. Dupont S, Morsut L, Aragona M, Enzo E, Giulitti S, Cordenonsi M, Zanconato F, Le Digabel J, Forcato M, Bicciato S, Elvassore N, Piccolo S. Role of YAP/TAZ in mechanotransduction. *Nature*. 2011 Jun 8;474(7350):179-83.
20. Calvo F, Ege N, Grande-Garcia A, Hooper S, Jenkins RP, Chaudhry SI, Harrington K, Williamson P, Moeendarbary E, Charras G, Sahai E. Mechanotransduction and YAP-dependent matrix remodelling is required for the generation and maintenance of cancer-associated fibroblasts. *Nat Cell Biol*. 2013 Jun;15(6):637-46.
21. Cotton JL, Li Q, Ma L, Park JS, Wang J, Ou J, Zhu LJ, Ip YT, Johnson RL, Mao J. YAP/TAZ and Hedgehog Coordinate Growth and Patterning in Gastrointestinal Mesenchyme. *Dev Cell*. 2017 Oct 9;43(1):35-47.e4.
22. Pipes GC, Creemers EE, Olson EN. The myocardin family of transcriptional coactivators: versatile regulators of cell growth, migration, and myogenesis. *Genes Dev*. 2006 Jun 15;20(12):1545-56.
23. Li X, Deng W, Nail CD, Bailey SK, Kraus MH, Ruppert JM, Lobo-Ruppert SM. Snail induction is an early response to Gli1 that determines the efficiency of epithelial transformation. *Oncogene*. 2006 Jan 26;25(4):609-21.
24. Erez N, Truitt M, Olson P, Arron ST, Hanahan D. Cancer-Associated Fibroblasts Are Activated in Incipient Neoplasia to Orchestrate Tumor-Promoting Inflammation in an NF-kappaB-Dependent Manner. *Cancer Cell*. 2010 Feb 17;17(2):135-47.
25. Sullivan NJ, Sasser AK, Axel AE, Vesuna F, Raman V, Ramirez N, Oberyszyn TM, Hall BM. Interleukin-6 induces an epithelial-mesenchymal transition phenotype in human breast cancer cells. *Oncogene*. 2009 Aug 20;28(33):2940-7.
26. Ebbing EA, van der Zalm AP, Steins A, Creemers A, Hermsen S, Rentenaar R, Klein M, Waasdorp C, Hooijer GKJ, Meijer SL, Krishnadath KK, Punt CJA, van Berge Henegouwen MI, Gisbertz SS, van Delden OM, Hulshof MCCM, Medema JP, van Laarhoven HWM, Bijlsma MF. Stromal-derived interleukin 6 drives epithelial-to-mesenchymal transition and therapy resistance in esophageal adenocarcinoma. *Proc Natl Acad Sci U S A*. 2019 Feb 5;116(6):2237-2242.
27. Cai J, Tang H, Xu L, Wang X, Yang C, Ruan S, Guo J, Hu S, Wang Z. Fibroblasts in omentum activated by tumor cells promote ovarian cancer growth, adhesion and invasiveness. *Carcinogenesis*. 2012 Jan;33(1):20-9.
28. Westerlund A, Hujanen E, Puistola U, Turpeenniemi-Hujanen T. Fibroblasts stimulate human ovarian cancer cell invasion and expression of 72-kDa gelatinase A (MMP-2). *Gynecol Oncol*. 1997 Oct;67(1):76-82.
29. Rinn JL, Bondre C, Gladstone HB, Brown PO, Chang HY. Anatomic demarcation by

positional variation in fibroblast gene expression programs. *PLoS Genet*. 2006 Jul;2(7):e119.
30. Göritz C, Dias DO, Tomilin N, Barbacid M, Shupliakov O, Frisén J. A pericyte origin of spinal cord scar tissue. *Science*. 2011 Jul 8;333(6039):238-42.
31. Hosaka K, Yang Y, Seki T, Fischer C, Dubey O, Fredlund E, Hartman J, Religa P, Morikawa H, Ishii Y, Sasahara M, Larsson O, Cossu G, Cao R, Lim S, Cao Y. Pericyte-fibroblast transition promotes tumor growth and metastasis. *Proc Natl Acad Sci U S A*. 2016 Sep 20;113(38):E5618-27.
32. Zeisberg EM, Potenta S, Xie L, Zeisberg M, Kalluri R. Discovery of endothelial to mesenchymal transition as a source for carcinoma-associated fibroblasts. *Cancer Res*. 2007 Nov 1;67(21):10123-8.
33. Driskell RR, Lichtenberger BM, Hoste E, Kretzschmar K, Simons BD, Charalambous M, Ferron SR, Herault Y, Pavlovic G, Ferguson-Smith AC, Watt FM. Distinct fibroblast lineages determine dermal architecture in skin development and repair. *Nature*. 2013 Dec 12;504(7479):277-281.
34. Lynch MD, Watt FM. Fibroblast heterogeneity: implications for human disease. *J Clin Invest*. 2018 Jan 2;128(1):26-35.
35. Rinkevich Y, Walmsley GG, Hu MS, Maan ZN, Newman AM, Drukker M, Januszyk M, Krampitz GW, Gurtner GC, Lorenz HP, Weissman IL, Longaker MT. Skin fibrosis. Identification and isolation of a dermal lineage with intrinsic fibrogenic potential. *Science*. 2015 Apr 17;348(6232):aaa2151.
36. Hu B, Castillo E, Harewood L, Ostano P, Reymond A, Dummer R, Raffoul W, Hoetzenecker W, Hofbauer GF, Dotto GP. Multifocal epithelial tumors and field cancerization from loss of mesenchymal CSL signaling. *Cell*. 2012 Jun 8;149(6):1207-20.
37. Procopio MG, Laszlo C, Al Labban D, Kim DE, Bordignon P, Jo SH, Goruppi S, Menietti E, Ostano P, Ala U, Provero P, Hoetzenecker W, Neel V, Kilarski WW, Swartz MA, Brisken C, Lefort K, Dotto GP. Combined CSL and p53 downregulation promotes cancer-associated fibroblast activation. *Nat Cell Biol*. 2015 Sep;17(9):1193-204.
38. Barron L, Gharib SA, Duffield JS. Lung Pericytes and Resident Fibroblasts: Busy Multitaskers. *Am J Pathol*. 2016 Oct;186(10):2519-31.
39. Rock JR, Barkauskas CE, Cronce MJ, Xue Y, Harris JR, Liang J, Noble PW, Hogan BL. Multiple stromal populations contribute to pulmonary fibrosis without evidence for epithelial to mesenchymal transition. *Proc Natl Acad Sci U S A*. 2011 Dec 27;108(52):E1475-83.
40. Sousa AM, Liu T, Guevara O, Stevens J, Fanburg BL, Gaestel M, Toksoz D, Kayyali US. Smooth muscle alpha-actin expression and myofibroblast differentiation by TGFbeta are dependent upon MK2. *J Cell Biochem*. 2007 Apr 15;100(6):1581-92.
41. Hashimoto N, Jin H, Liu T, Chensue SW, Phan SH. Bone marrow-derived progenitor cells in pulmonary fibrosis. *J Clin Invest*. 2004 Jan;113(2):243-52.
42. Kim KK, Kugler MC, Wolters PJ, Robillard L, Galvez MG, Brumwell AN, Sheppard D, Chapman HA. Alveolar epithelial cell mesenchymal transition develops in vivo during pulmonary fibrosis and is regulated by the extracellular matrix. *Proc Natl Acad Sci U S A*. 2006 Aug 29;103(35):13180-5.
43. Cruz-Bermúdez A, Laza-Briviesca R, Vicente-Blanco RJ, García-Grande A,

Coronado MJ, Laine-Menéndez S, Alfaro C, Sanchez JC, Franco F, Calvo V, Romero A, Martin-Acosta P, Salas C, Garcia JM, Provencio M. Cancer-associated fibroblasts modify lung cancer metabolism involving ROS and TGF-β signaling. *Free Radic Biol Med*. 2019 Jan;130:163-173.
44. Zhang W, Bouchard G, Yu A, Shafiq M, Jamali M, Shrager JB, Ayers K, Bakr S, Gentles AJ, Diehn M, Quon A, West RB, Nair V, van de Rijn M, Napel S, Plevritis SK. GFPT2-Expressing Cancer-Associated Fibroblasts Mediate Metabolic Reprogramming in Human Lung Adenocarcinoma. *Cancer Res*. 2018 Jul 1;78(13):3445-3457.
45. Rinkevich Y, Mori T, Sahoo D, Xu PX, Bermingham JR Jr, Weissman IL. Identification and prospective isolation of a mesothelial precursor lineage giving rise to smooth muscle cells and fibroblasts for mammalian internal organs, and their vasculature. *Nat Cell Biol*. 2012 Dec;14(12):1251-60.
46. Koliaraki V, Pallangyo CK, Greten FR, Kollias G. Mesenchymal Cells in Colon Cancer. *Gastroenterology*. 2017 Apr;152(5):964-979.
47. Powell DW, Pinchuk IV, Saada JI, Chen X, Mifflin RC. Mesenchymal cells of the intestinal lamina propria. *Annu Rev Physiol*. 2011;73:213-37.
48. Quante M, Tu SP, Tomita H, Gonda T, Wang SS, Takashi S, Baik GH, Shibata W, Diprete B, Betz KS, Friedman R, Varro A, Tycko B, Wang TC. Bone marrow-derived myofibroblasts contribute to the mesenchymal stem cell niche and promote tumor growth. *Cancer Cell*. 2011 Feb 15;19(2):257-72.
49. Theiss AL, Simmons JG, Jobin C, Lund PK. Tumor necrosis factor (TNF) alpha increases collagen accumulation and proliferation in intestinal myofibroblasts via TNF receptor 2. *J Biol Chem*. 2005 Oct 28;280(43):36099-109.
50. Gong Y, Scott E, Lu R, Xu Y, Oh WK, Yu Q. TIMP-1 promotes accumulation of cancer associated fibroblasts and cancer progression. *PLoS One*. 2013 Oct 15;8(10):e77366.
51. Qiu W, Hu M, Sridhar A, Opeskin K, Fox S, Shipitsin M, Trivett M, Thompson ER, Ramakrishna M, Gorringe KL, Polyak K, Haviv I, Campbell IG. No evidence of clonal somatic genetic alterations in cancer-associated fibroblasts from human breast and ovarian carcinomas. *Nat Genet*. 2008 May;40(5):650-5.
52. Morsing M, Klitgaard MC, Jafari A, Villadsen R, Kassem M, Petersen OW, Rønnov-Jessen L. Evidence of two distinct functionally specialized fibroblast lineages in breast stroma. *Breast Cancer Res*. 2016 Nov 3;18(1):108.
53. Howard BA, Gusterson BA. Human breast development. J Mammary Gland Biol *Neoplasia*. 2000 Apr;5(2):119-37.
54. Howard BA, Lu P. Stromal regulation of embryonic and postnatal mammary epithelial development and differentiation. *Semin Cell Dev Biol*. 2014 Jan-Feb;25-26:43-51.
55. Costa A, Kieffer Y, Scholer-Dahirel A, Pelon F, Bourachot B, Cardon M, Sirven P, Magagna I, Fuhrmann L, Bernard C, Bonneau C, Kondratova M, Kuperstein I, Zinovyev A, Givel AM, Parrini MC, Soumelis V, Vincent-Salomon A, Mechta-Grigoriou F. Fibroblast Heterogeneity and Immunosuppressive Environment in Human Breast Cancer. *Cancer Cell*. 2018 Mar 12;33(3):463-479.e10.
56. Su S, Chen J, Yao H, Liu J, Yu S, Lao L, Wang M, Luo M, Xing Y, Chen F, Huang D, Zhao J, Yang L, Liao D, Su F, Li M, Liu Q, Song E. CD10(+)GPR77(+) Cancer-Associated Fibroblasts Promote Cancer Formation and Chemoresistance by Sustaining Cancer Stemness. *Cell*. 2018 Feb 8;172(4):841-856.e16.

57. Toullec A, Gerald D, Despouy G, Bourachot B, Cardon M, Lefort S, Richardson M, Rigaill G, Parrini MC, Lucchesi C, Bellanger D, Stern MH, Dubois T, Sastre-Garau X, Delattre O, Vincent-Salomon A, Mechta-Grigoriou F. Oxidative stress promotes myofibroblast differentiation and tumour spreading. *EMBO Mol Med*. 2010 Jun;2(6):211-30.
58. Houthuijzen JM, Jonkers J. Cancer-associated fibroblasts as key regulators of the breast cancer tumor microenvironment. *Cancer Metastasis Rev*. 2018 Dec;37(4):577-597.
59. Bachem MG, Schneider E, Gross H, Weidenbach H, Schmid RM, Menke A, Siech M, Beger H, Grünert A, Adler G. Identification, culture, and characterization of pancreatic stellate cells in rats and humans. *Gastroenterology*. 1998 Aug;115(2):421-32.
60. Schneider E, Schmid-Kotsas A, Zhao J, Weidenbach H, Schmid RM, Menke A, Adler G, Waltenberger J, Grünert A, Bachem MG. Identification of mediators stimulating proliferation and matrix synthesis of rat pancreatic stellate cells. *Am J Physiol Cell Physiol*. 2001 Aug;281(2):C532-43.
61. Omary MB, Lugea A, Lowe AW, Pandol SJ. The pancreatic stellate cell: a star on the rise in pancreatic diseases. *J Clin Invest*. 2007 Jan;117(1):50-9.
62. Bachem MG, Schünemann M, Ramadani M, Siech M, Beger H, Buck A, Zhou S, Schmid-Kotsas A, Adler G. Pancreatic carcinoma cells induce fibrosis by stimulating proliferation and matrix synthesis of stellate cells. *Gastroenterology*. 2005 Apr;128(4):907-21.
63. Ko SY, Barengo N, Ladanyi A, Lee JS, Marini F, Lengyel E, Naora H. HOXA9 promotes ovarian cancer growth by stimulating cancer-associated fibroblasts. *J Clin Invest*. 2012 Oct;122(10):3603-17.
64. Öhlund D, Handly-Santana A, Biffi G, Elyada E, Almeida AS, Ponz-Sarvise M, Corbo V, Oni TE, Hearn SA, Lee EJ, Chio II, Hwang CI, Tiriac H, Baker LA, Engle DD, Feig C, Kultti A, Egeblad M, Fearon DT, Crawford JM, Clevers H, Park Y, Tuveson DA. Distinct populations of inflammatory fibroblasts and myofibroblasts in pancreatic cancer. *J Exp Med*. 2017 Mar 6;214(3):579-596.
65. Neuschwander-Tetri BA, Bridle KR, Wells LD, Marcu M, Ramm GA. Repetitive acute pancreatic injury in the mouse induces procollagen alpha1(I) expression colocalized to pancreatic stellate cells. *Lab Invest*. 2000 Feb;80(2):143-50.
66. Binkley CE, Zhang L, Greenson JK, Giordano TJ, Kuick R, Misek D, Hanash S, Logsdon CD, Simeone DM. The molecular basis of pancreatic fibrosis: common stromal gene expression in chronic pancreatitis and pancreatic adenocarcinoma. *Pancreas*. 2004 Nov;29(4):254-63.
67. Elyada E, Bolisetty M, Laise P, Flynn WF, Courtois ET, Burkhart RA, Teinor JA, Belleau P, Biffi G, Lucito MS, Sivajothi S, Armstrong TD, Engle DD, Yu KH, Hao Y, Wolfgang CL, Park Y, Preall J, Jaffee EM, Califano A, Robson P, Tuveson DA. Cross-Species Single-Cell Analysis of Pancreatic Ductal Adenocarcinoma Reveals Antigen-Presenting Cancer-Associated Fibroblasts. *Cancer Discov*. 2019 Aug;9(8):1102-1123.
68. King D, Anatomy, School of Medicine, Southern Illinois University, http://www.siumed.edu/~dking2/index.htm
69. Wang W, Kryczek I, Dostál L, Lin H, Tan L, Zhao L, Lu F, Wei S, Maj T, Peng D, He G, Vatan L, Szeliga W, Kuick R, Kotarski J, Tarkowski R, Dou Y, Rattan R, Munkarah A, Liu JR, Zou W. Effector T Cells Abrogate Stroma-Mediated

Chemoresistance in Ovarian Cancer. *Cell*. 2016 May 19;165(5):1092-1105.
70. Curtis M, Kenny HA, Ashcroft B, Mukherjee A, Johnson A, Zhang Y, Helou Y, Batlle R, Liu X, Gutierrez N, Gao X, Yamada SD, Lastra R, Montag A, Ahsan N, Locasale JW, Salomon AR, Nebreda AR, Lengyel E. Fibroblasts Mobilize Tumor Cell Glycogen to Promote Proliferation and Metastasis. *Cell Metab*. 2019 Jan 8;29(1):141-155.e9.
71. Mitra AK, Zillhardt M, Hua Y, Tiwari P, Murmann AE, Peter ME, Lengyel E. MicroRNAs reprogram normal fibroblasts into cancer-associated fibroblasts in ovarian cancer. *Cancer Discov*. 2012 Dec;2(12):1100-8.
72. Dirat B, Bochet L, Dabek M, Daviaud D, Dauvillier S, Majed B, Wang YY, Meulle A, Salles B, Le Gonidec S, Garrido I, Escourrou G, Valet P, Muller C. Cancer-associated adipocytes exhibit an activated phenotype and contribute to breast cancer invasion. *Cancer Res*. 2011 Apr 1;71(7):2455-65.
73. Bochet L, Lehuédé C, Dauvillier S, Wang YY, Dirat B, Laurent V, Dray C, Guiet R, Maridonneau-Parini I, Le Gonidec S, Couderc B, Escourrou G, Valet P, Muller C. Adipocyte-derived fibroblasts promote tumor progression and contribute to the desmoplastic reaction in breast cancer. *Cancer Res*. 2013 Sep 15;73(18):5657-68.
74. Gustafson B, Smith U. Activation of canonical wingless-type MMTV integration site family (Wnt) signaling in mature adipocytes increases beta-catenin levels and leads to cell dedifferentiation and insulin resistance. *J Biol Chem*. 2010 Apr 30;285(18):14031-41.
75. Özdemir BC, Pentcheva-Hoang T, Carstens JL, Zheng X, Wu CC, Simpson TR, Laklai H, Sugimoto H, Kahlert C, Novitskiy SV, De Jesus-Acosta A, Sharma P, Heidari P, Mahmood U, Chin L, Moses HL, Weaver VM, Maitra A, Allison JP, LeBleu VS, Kalluri R. Depletion of carcinoma-associated fibroblasts and fibrosis induces immunosuppression and accelerates pancreas cancer with reduced survival. *Cancer Cell*. 2014 Jun 16;25(6):719-34.
76. Bhowmick NA, Chytil A, Plieth D, Gorska AE, Dumont N, Shappell S, Washington MK, Neilson EG, Moses HL. TGF-beta signaling in fibroblasts modulates the oncogenic potential of adjacent epithelia. *Science*. 2004 Feb 6;303(5659):848-51.
77. Katajisto P, Vaahtomeri K, Ekman N, Ventelä E, Ristimäki A, Bardeesy N, Feil R, DePinho RA, Mäkelä TP. LKB1 signaling in mesenchymal cells required for suppression of gastrointestinal polyposis. *Nat Genet*. 2008 Apr;40(4):455-9.
78. Tanwar PS, Kaneko-Tarui T, Zhang L, Tanaka Y, Crum CP, Teixeira JM. Stromal liver kinase B1 [STK11] signaling loss induces oviductal adenomas and endometrial cancer by activating mammalian Target of Rapamycin Complex 1. *PLoS Genet*. 2012;8(8):e1002906.
79. Contreras CM, Gurumurthy S, Haynie JM, Shirley LJ, Akbay EA, Wingo SN, Schorge JO, Broaddus RR, Wong KK, Bardeesy N, Castrillon DH. Loss of Lkb1 provokes highly invasive endometrial adenocarcinomas. *Cancer Res*. 2008 Feb 1;68(3):759-66.
80. Chen H, Detmer SA, Ewald AJ, Griffin EE, Fraser SE, Chan DC. Mitofusins Mfn1 and Mfn2 coordinately regulate mitochondrial fusion and are essential for embryonic development. *J Cell Biol*. 2003 Jan 20;160(2):189-200.
81. Koshiba T, Detmer SA, Kaiser JT, Chen H, McCaffery JM, Chan DC. Structural basis of mitochondrial tethering by mitofusin complexes. *Science*. 2004 Aug 6;305(5685):858-62.
82. Song Z, Chen H, Fiket M, Alexander C, Chan DC. OPA1 processing controls

mitochondrial fusion and is regulated by mRNA splicing, membrane potential, and Yme1L. *J Cell Biol*. 2007 Aug 27;178(5):749-55.
83. Ishihara N, Fujita Y, Oka T, Mihara K. Regulation of mitochondrial morphology through proteolytic cleavage of OPA1. *EMBO J*. 2006 Jul 12;25(13):2966-77.
84. Korobova F, Ramabhadran V, Higgs HN. An actin-dependent step in mitochondrial fission mediated by the ER-associated formin INF2. *Science*. 2013 Jan 25;339(6118):464-7.
85. Lee JE, Westrate LM, Wu H, Page C, Voeltz GK. Multiple dynamin family members collaborate to drive mitochondrial division. *Nature*. 2016 Dec 1;540(7631):139-143.
86. Zhang D, Wang Y, Shi Z, Liu J, Sun P, Hou X, Zhang J, Zhao S, Zhou BP, Mi J. Metabolic reprogramming of cancer-associated fibroblasts by IDH3α downregulation. *Cell Rep*. 2015 Mar 3;10(8):1335-48.
87. Fiaschi T, Marini A, Giannoni E, Taddei ML, Gandellini P, De Donatis A, Lanciotti M, Serni S, Cirri P, Chiarugi P. Reciprocal metabolic reprogramming through lactate shuttle coordinately influences tumor-stroma interplay. *Cancer Res*. 2012 Oct 1;72(19):5130-40.
88. Ippolito L, Morandi A, Taddei ML, Parri M, Comito G, Iscaro A, Raspollini MR, Magherini F, Rapizzi E, Masquelier J, Muccioli GG, Sonveaux P, Chiarugi P, Giannoni E. Cancer-associated fibroblasts promote prostate cancer malignancy via metabolic rewiring and mitochondrial transfer. *Oncogene*. 2019 Jul;38(27):5339-5355.
89. Yan W, Wu X, Zhou W, Fong MY, Cao M, Liu J, Liu X, Chen CH, Fadare O, Pizzo DP, Wu J, Liu L, Liu X, Chin AR, Ren X, Chen Y, Locasale JW, Wang SE. Cancer-cell-secreted exosomal miR-105 promotes tumour growth through the MYC-dependent metabolic reprogramming of stromal cells. *Nat Cell Biol*. 2018 May;20(5):597-609.
90. Tejerina S, De Pauw A, Vankoningsloo S, Houbion A, Renard P, De Longueville F, Raes M, Arnould T. Mild mitochondrial uncoupling induces 3T3-L1 adipocyte de-differentiation by a PPARgamma-independent mechanism, whereas TNFalpha-induced de-differentiation is PPARgamma dependent. *J Cell Sci*. 2009 Jan 1;122(Pt 1):145-55.
91. Bertero T, Oldham WM, Grasset EM, Bourget I, Boulter E, Pisano S, Hofman P, Bellvert F, Meneguzzi G, Bulavin DV, Estrach S, Feral CC, Chan SY, Bozec A, Gaggioli C. Tumor-Stroma Mechanics Coordinate Amino Acid Availability to Sustain Tumor Growth and Malignancy. *Cell Metab*. 2019 Jan 8;29(1):124-140.e10.
92. Hemminki A, Markie D, Tomlinson I, Avizienyte E, Roth S, Loukola A, Bignell G, Warren W, Aminoff M, Höglund P, Järvinen H, Kristo P, Pelin K, Ridanpää M, Salovaara R, Toro T, Bodmer W, Olschwang S, Olsen AS, Stratton MR, de la Chapelle A, Aaltonen LA. A serine/threonine kinase gene defective in Peutz-Jeghers syndrome. *Nature*. 1998 Jan 8;391(6663):184-7.
93. Lo B, Strasser G, Sagolla M, Austin CD, Junttila M, Mellman I. Lkb1 regulates organogenesis and early oncogenesis along AMPK-dependent and -independent pathways. *J Cell Biol*. 2012 Dec 24;199(7):1117-30.
94. Nakada D, Saunders TL, Morrison SJ. Lkb1 regulates cell cycle and energy metabolism in haematopoietic stem cells. *Nature*. 2010 Dec 2;468(7324):653-8.
95. Gan B, Hu J, Jiang S, Liu Y, Sahin E, Zhuang L, Fletcher-Sananikone E, Colla S, Wang YA, Chin L, Depinho RA. Lkb1 regulates quiescence and metabolic homeostasis of haematopoietic stem cells. *Nature*. 2010 Dec 2;468(7324):701-4.

96. Gurumurthy S, Xie SZ, Alagesan B, Kim J, Yusuf RZ, Saez B, Tzatsos A, Ozsolak F, Milos P, Ferrari F, Park PJ, Shirihai OS, Scadden DT, Bardeesy N. The Lkb1 metabolic sensor maintains haematopoietic stem cell survival. *Nature*. 2010 Dec 2;468(7324):659-63.
97. Yang K, Blanco DB, Neale G, Vogel P, Avila J, Clish CB, Wu C, Shrestha S, Rankin S, Long L, Kc A, Chi H. Homeostatic control of metabolic and functional fitness of T(reg) cells by LKB1 signalling. *Nature*. 2017 Aug 31;548(7669):602-606.
98. Lizcano JM, Göransson O, Toth R, Deak M, Morrice NA, Boudeau J, Hawley SA, Udd L, Mäkelä TP, Hardie DG, Alessi DR. LKB1 is a master kinase that activates 13 kinases of the AMPK subfamily, including MARK/PAR-1. *EMBO J*. 2004 Feb 25;23(4):833-43.
99. Hawley SA, Boudeau J, Reid JL, Mustard KJ, Udd L, Mäkelä TP, Alessi DR, Hardie DG. Complexes between the LKB1 tumor suppressor, STRAD alpha/beta and MO25 alpha/beta are upstream kinases in the AMP-activated protein kinase cascade. *J Biol*. 2003;2(4):28.
100. Shaw RJ, Kosmatka M, Bardeesy N, Hurley RL, Witters LA, DePinho RA, Cantley LC. The tumor suppressor LKB1 kinase directly activates AMP-activated kinase and regulates apoptosis in response to energy stress. *Proc Natl Acad Sci U S A*. 2004 Mar 9;101(10):3329-35.
101. Jaleel M, McBride A, Lizcano JM, Deak M, Toth R, Morrice NA, Alessi DR. Identification of the sucrose non-fermenting related kinase SNRK, as a novel LKB1 substrate. *FEBS Lett*. 2005 Feb 28;579(6):1417-23.
102. Shaw RJ, Bardeesy N, Manning BD, Lopez L, Kosmatka M, DePinho RA, Cantley LC. The LKB1 tumor suppressor negatively regulates mTOR signaling. *Cancer Cell*. 2004 Jul;6(1):91-9.
103. Karuman P, Gozani O, Odze RD, Zhou XC, Zhu H, Shaw R, Brien TP, Bozzuto CD, Ooi D, Cantley LC, Yuan J. The Peutz-Jegher gene product LKB1 is a mediator of p53-dependent cell death. *Mol Cell*. 2001 Jun;7(6):1307-19.
104. Morén A, Raja E, Heldin CH, Moustakas A. Negative regulation of TGFβ signaling by the kinase LKB1 and the scaffolding protein LIP1. *J Biol Chem*. 2011 Jan 7;286(1):341-53.
105. Ollila S, Domènech-Moreno E, Laajanen K, Wong IP, Tripathi S, Pentinmikko N, Gao Y, Yan Y, Niemelä EH, Wang TC, Viollet B, Leone G, Katajisto P, Vaahtomeri K, Mäkelä TP. Stromal Lkb1 deficiency leads to gastrointestinal tumorigenesis involving the IL-11-JAK/STAT3 pathway. *J Clin Invest*. 2018 Jan 2;128(1):402-414.
106. Vaahtomeri K, Ventelä E, Laajanen K, Katajisto P, Wipff PJ, Hinz B, Vallenius T, Tiainen M, Mäkelä TP. Lkb1 is required for TGFbeta-mediated myofibroblast differentiation. *J Cell Sci*. 2008 Nov 1;121(Pt 21):3531-40.
107. Ring KL, Garcia C, Thomas MH, Modesitt SC. Current and future role of genetic screening in gynecologic malignancies. *Am J Obstet Gynecol*. 2017 Nov;217(5):512-521.
108. Giardiello FM, Welsh SB, Hamilton SR, Offerhaus GJ, Gittelsohn AM, Booker SV, Krush AJ, Yardley JH, Luk GD. Increased risk of cancer in the Peutz-Jeghers syndrome. *N Engl J Med*. 1987 Jun 11;316(24):1511-4.
109. Poffenberger MC, Metcalfe-Roach A, Aguilar E, Chen J, Hsu BE, Wong AH, Johnson RM, Flynn B, Samborska B, Ma EH, Gravel SP, Tonelli L, Devorkin L, Kim P, Hall A, Izreig S, Loginicheva E, Beauchemin N, Siegel PM, Artyomov MN, Lum JJ,

Zogopoulos G, Blagih J, Jones RG. LKB1 deficiency in T cells promotes the development of gastrointestinal polyposis. *Science*. 2018 Jul 27;361(6400):406-411.
110. He N, Fan W, Henriquez B, Yu RT, Atkins AR, Liddle C, Zheng Y, Downes M, Evans RM. Metabolic control of regulatory T cell (Treg) survival and function by Lkb1. *Proc Natl Acad Sci U S A*. 2017 Nov 21;114(47):12542-12547.
111. Xu HG, Zhai YX, Chen J, Lu Y, Wang JW, Quan CS, Zhao RX, Xiao X, He Q, Werle KD, Kim HG, Lopez R, Cui R, Liang J, Li YL, Xu ZX. LKB1 reduces ROS-mediated cell damage via activation of p38. *Oncogene*. 2015 Jul;34(29):3848-59.
112. Hawley SA, Pan DA, Mustard KJ, Ross L, Bain J, Edelman AM, Frenguelli BG, Hardie DG. Calmodulin-dependent protein kinase kinase-beta is an alternative upstream kinase for AMP-activated protein kinase. *Cell Metab*. 2005 Jul;2(1):9-19.
113. Jones RG, Plas DR, Kubek S, Buzzai M, Mu J, Xu Y, Birnbaum MJ, Thompson CB. AMP-activated protein kinase induces a p53-dependent metabolic checkpoint. *Mol Cell*. 2005 Apr 29;18(3):283-93.
114. Eichner LJ, Brun SN, Herzig S, Young NP, Curtis SD, Shackelford DB, Shokhirev MN, Leblanc M, Vera LI, Hutchins A, Ross DS, Shaw RJ, Svensson RU. Genetic Analysis Reveals AMPK Is Required to Support Tumor Growth in Murine Kras-Dependent Lung Cancer Models. *Cell Metab*. 2019 Feb 5;29(2):285-302.e7.
115. Budanov AV, Karin M. p53 target genes sestrin1 and sestrin2 connect genotoxic stress and mTOR signaling. *Cell*. 2008 Aug 8;134(3):451-60.
116. Hardie DG, Scott JW, Pan DA, Hudson ER. Management of cellular energy by the AMP-activated protein kinase system. *FEBS Lett*. 2003 Jul 3;546(1):113-20.
117. Suzuki A, Hirata M, Kamimura K, Maniwa R, Yamanaka T, Mizuno K, Kishikawa M, Hirose H, Amano Y, Izumi N, Miwa Y, Ohno S. aPKC acts upstream of PAR-1b in both the establishment and maintenance of mammalian epithelial polarity. *Curr Biol*. 2004 Aug 24;14(16):1425-35.
118. Hurov JB, Watkins JL, Piwnica-Worms H. Atypical PKC phosphorylates PAR-1 kinases to regulate localization and activity. *Curr Biol*. 2004 Apr 20;14(8):736-41.
119. Sun TQ, Lu B, Feng JJ, Reinhard C, Jan YN, Fantl WJ, Williams LT. PAR-1 is a Dishevelled-associated kinase and a positive regulator of Wnt signalling. *Nat Cell Biol*. 2001 Jul;3(7):628-36.
120. Ossipova O, Dhawan S, Sokol S, Green JB. Distinct PAR-1 proteins function in different branches of Wnt signaling during vertebrate development. *Dev Cell*. 2005 Jun;8(6):829-41.
121. Zhang X, Zhu J, Yang GY, Wang QJ, Qian L, Chen YM, Chen F, Tao Y, Hu HS, Wang T, Luo ZG. Dishevelled promotes axon differentiation by regulating atypical protein kinase C. *Nat Cell Biol*. 2007 Jul;9(7):743-54.
122. Kojima Y, Miyoshi H, Clevers HC, Oshima M, Aoki M, Taketo MM. Suppression of tubulin polymerization by the LKB1-microtubule-associated protein/microtubule affinity-regulating kinase signaling. *J Biol Chem*. 2007 Aug 10;282(32):23532-40.
123. Sandí MJ, Marshall CB, Balan M, Coyaud É, Zhou M, Monson DM, Ishiyama N, Chandrakumar AA, La Rose J, Couzens AL, Gingras AC, Raught B, Xu W, Ikura M, Morrison DK, Rottapel R. MARK3-mediated phosphorylation of ARHGEF2 couples microtubules to the actin cytoskeleton to establish cell polarity. *Sci Signal*. 2017 Oct 31;10(503). pii: eaan3286.
124. Mohseni M, Sun J, Lau A, Curtis S, Goldsmith J, Fox VL, Wei C, Frazier M,

Samson O, Wong KK, Kim C, Camargo FD. A genetic screen identifies an LKB1-MARK signalling axis controlling the Hippo-YAP pathway. *Nat Cell Biol*. 2014 Jan;16(1):108-17.
125. Nguyen HB, Babcock JT, Wells CD, Quilliam LA. LKB1 tumor suppressor regulates AMP kinase/mTOR-independent cell growth and proliferation via the phosphorylation of Yap. *Oncogene*. 2013 Aug 29;32(35):4100-9.
126. Lennerz JK, Hurov JB, White LS, Lewandowski KT, Prior JL, Planer GJ, Gereau RW 4th, Piwnica-Worms D, Schmidt RE, Piwnica-Worms H. Loss of Par-1a/MARK3/C-TAK1 kinase leads to reduced adiposity, resistance to hepatic steatosis, and defective gluconeogenesis. *Mol Cell Biol*. 2010 Nov;30(21):5043-56.
127. Sun C, Tian L, Nie J, Zhang H, Han X, Shi Y. Inactivation of MARK4, an AMP-activated protein kinase (AMPK)-related kinase leads to insulin hypersensitivity and resistance to diet-induced obesity. *J Biol Chem*. 2012 Nov 2;287(45):38305-15.
128. Hurov JB, Huang M, White LS, Lennerz J, Choi CS, Cho YR, Kim HJ, Prior JL, Piwnica-Worms D, Cantley LC, Kim JK, Shulman GI, Piwnica-Worms H. Loss of the Par-1b/MARK2 polarity kinase leads to increased metabolic rate, decreased adiposity, and insulin hypersensitivity in vivo. *Proc Natl Acad Sci U S A*. 2007 Mar 27;104(13):5680-5.
129. Suzuki A, Kusakai G, Kishimoto A, Lu J, Ogura T, Lavin MF, Esumi H. Identification of a novel protein kinase mediating Akt survival signaling to the ATM protein. *J Biol Chem*. 2003 Jan 3;278(1):48-53.
130. Suzuki A, Ogura T, Esumi H. NDR2 acts as the upstream kinase of ARK5 during insulin-like growth factor-1 signaling. *J Biol Chem*. 2006 May 19;281(20):13915-21.
131. Kolliopoulos C, Raja E, Razmara M, Heldin P, Heldin CH, Moustakas A, van der Heide LP. Transforming growth factor β (TGFβ) induces NUAK kinase expression to fine-tune its signaling output. *J Biol Chem*. 2019 Mar 15;294(11):4119-4136.
132. Yuan WC, Pepe-Mooney B, Galli GG, Dill MT, Huang HT, Hao M, Wang Y, Liang H, Calogero RA, Camargo FD. NUAK2 is a critical YAP target in liver cancer. *Nat Commun*. 2018 Nov 16;9(1):4834.
133. Zagórska A, Deak M, Campbell DG, Banerjee S, Hirano M, Aizawa S, Prescott AR, Alessi DR. New roles for the LKB1-NUAK pathway in controlling myosin phosphatase complexes and cell adhesion. *Sci Signal*. 2010 Mar 30;3(115):ra25.
134. Yamamoto H, Takashima S, Shintani Y, Yamazaki S, Seguchi O, Nakano A, Higo S, Kato H, Liao Y, Asano Y, Minamino T, Matsumura Y, Takeda H, Kitakaze M. Identification of a novel substrate for TNFalpha-induced kinase NUAK2. *Biochem Biophys Res Commun*. 2008 Jan 18;365(3):541-7.
135. Monteverde T, Tait-Mulder J, Hedley A, Knight JR, Sansom OJ, Murphy DJ. Calcium signalling links MYC to NUAK1. Oncogene. 2018 Feb 22;37(8):982-992.
136. Hou X, Liu JE, Liu W, Liu CY, Liu ZY, Sun ZY. A new role of NUAK1: directly phosphorylating p53 and regulating cell proliferation. *Oncogene*. 2011 Jun 30;30(26):2933-42.
137. Humbert N, Navaratnam N, Augert A, Da Costa M, Martien S, Wang J, Martinez D, Abbadie C, Carling D, de Launoit Y, Gil J, Bernard D. Regulation of ploidy and senescence by the AMPK-related kinase NUAK1. *EMBO J*. 2010 Jan 20;29(2):376-86.
138. Liu L, Ulbrich J, Müller J, Wüstefeld T, Aeberhard L, Kress TR, Muthalagu N, Rycak L, Rudalska R, Moll R, Kempa S, Zender L, Eilers M, Murphy DJ. Deregulated

MYC expression induces dependence upon AMPK-related kinase 5. *Nature*. 2012 Mar 28;483(7391):608-12.
139. Courchet J, Lewis TL Jr, Lee S, Courchet V, Liou DY, Aizawa S, Polleux F. Terminal axon branching is regulated by the LKB1-NUAK1 kinase pathway via presynaptic mitochondrial capture. *Cell*. 2013 Jun 20;153(7):1510-25.
140. Berdeaux R, Goebel N, Banaszynski L, Takemori H, Wandless T, Shelton GD, Montminy M. SIK1 is a class II HDAC kinase that promotes survival of skeletal myocytes. *Nat Med*. 2007 May;13(5):597-603.
141. Tarumoto Y, Lu B, Somerville TDD, Huang YH, Milazzo JP, Wu XS, Klingbeil O, El Demerdash O, Shi J, Vakoc CR. LKB1, Salt-Inducible Kinases, and MEF2C Are Linked Dependencies in Acute Myeloid Leukemia. *Mol Cell*. 2018 Mar 15;69(6):1017-1027.e6.
142. Screaton RA, Conkright MD, Katoh Y, Best JL, Canettieri G, Jeffries S, Guzman E, Niessen S, Yates JR 3rd, Takemori H, Okamoto M, Montminy M. The CREB coactivator TORC2 functions as a calcium- and cAMP-sensitive coincidence detector. *Cell*. 2004 Oct 1;119(1):61-74.
143. Koo SH, Flechner L, Qi L, Zhang X, Screaton RA, Jeffries S, Hedrick S, Xu W, Boussouar F, Brindle P, Takemori H, Montminy M. The CREB coactivator TORC2 is a key regulator of fasting glucose metabolism. *Nature*. 2005 Oct 20;437(7062):1109-11.
144. Kowanetz M, Lönn P, Vanlandewijck M, Kowanetz K, Heldin CH, Moustakas A. TGFbeta induces SIK to negatively regulate type I receptor kinase signaling. *J Cell Biol*. 2008 Aug 25;182(4):655-62.
145. Honda T, Fujiyama T, Miyoshi C, Ikkyu A, Hotta-Hirashima N, Kanno S, Mizuno S, Sugiyama F, Takahashi S, Funato H, Yanagisawa M. A single phosphorylation site of SIK3 regulates daily sleep amounts and sleep need in mice. *Proc Natl Acad Sci U S A*. 2018 Oct 9;115(41):10458-10463.
146. Funato H, Miyoshi C, Fujiyama T, Kanda T, Sato M, Wang Z, Ma J, Nakane S, Tomita J, Ikkyu A, Kakizaki M, Hotta-Hirashima N, Kanno S, Komiya H, Asano F, Honda T, Kim SJ, Harano K, Muramoto H, Yonezawa T, Mizuno S, Miyazaki S, Connor L, Kumar V, Miura I, Suzuki T, Watanabe A, Abe M, Sugiyama F, Takahashi S, Sakimura K, Hayashi Y, Liu Q, Kume K, Wakana S, Takahashi JS, Yanagisawa M. Forward-genetics analysis of sleep in randomly mutagenized mice. *Nature*. 2016 Nov 17;539(7629):378-383.
147. Miranda F, Mannion D, Liu S, Zheng Y, Mangala LS, Redondo C, Herrero-Gonzalez S, Xu R, Taylor C, Chedom DF, Karaminejadranjbar M, Albukhari A, Jiang D, Pradeep S, Rodriguez-Aguayo C, Lopez-Berestein G, Salah E, Abdul Azeez KR, Elkins JM, Campo L, Myers KA, Klotz D, Bivona S, Dhar S, Bast RC Jr, Saya H, Choi HG, Gray NS, Fischer R, Kessler BM, Yau C, Sood AK, Motohara T, Knapp S, Ahmed AA. Salt-Inducible Kinase 2 Couples Ovarian Cancer Cell Metabolism with Survival at the Adipocyte-Rich Metastatic Niche. *Cancer Cell*. 2016 Aug 8;30(2):273-289.
148. Barnes AP, Lilley BN, Pan YA, Plummer LJ, Powell AW, Raines AN, Sanes JR, Polleux F. LKB1 and SAD kinases define a pathway required for the polarization of cortical neurons. *Cell*. 2007 May 4;129(3):549-63.
149. Kishi M, Pan YA, Crump JG, Sanes JR. Mammalian SAD kinases are required for neuronal polarization. *Science*. 2005 Feb 11;307(5711):929-32.
150. Lilley BN, Pan YA, Sanes JR. SAD kinases sculpt axonal arbors of sensory

neurons through long- and short-term responses to neurotrophin signals. *Neuron*. 2013 Jul 10;79(1):39-53.
151. Inoue E, Mochida S, Takagi H, Higa S, Deguchi-Tawarada M, Takao-Rikitsu E, Inoue M, Yao I, Takeuchi K, Kitajima I, Setou M, Ohtsuka T, Takai Y. SAD: a presynaptic kinase associated with synaptic vesicles and the active zone cytomatrix that regulates neurotransmitter release. *Neuron*. 2006 Apr 20;50(2):261-75.
152. Nie J, Liu X, Lilley BN, Zhang H, Pan YA, Kimball SR, Zhang J, Zhang W, Wang L, Jefferson LS, Sanes JR, Han X, Shi Y. SAD-A kinase controls islet β-cell size and function as a mediator of mTORC1 signaling. *Proc Natl Acad Sci U S A*. 2013 Aug 20;110(34):13857-62.
153. Chun CZ, Kaur S, Samant GV, Wang L, Pramanik K, Garnaas MK, Li K, Field L, Mukhopadhyay D, Ramchandran R. Snrk-1 is involved in multiple steps of angioblast development and acts via notch signaling pathway in artery-vein specification in vertebrates. *Blood*. 2009 Jan 29;113(5):1192-9.
154. Li Y, Nie Y, Helou Y, Ding G, Feng B, Xu G, Salomon A, Xu H. Identification of sucrose non-fermenting-related kinase (SNRK) as a suppressor of adipocyte inflammation. *Diabetes*. 2013 Jul;62(7):2396-409.
155. Kim J, Coffey DM, Creighton CJ, Yu Z, Hawkins SM, Matzuk MM. High-grade serous ovarian cancer arises from fallopian tube in a mouse model. *Proc Natl Acad Sci U S A*. 2012 Mar 6;109(10):3921-6.
156. Zhang S, Dolgalev I, Zhang T, Ran H, Levine DA, Neel BG. Both fallopian tube and ovarian surface epithelium are cells-of-origin for high-grade serous ovarian carcinoma. *Nat Commun*. 2019 Nov 26;10(1):5367.
157. Singhal PK, Sassi S, Lan L, Au P, Halvorsen SC, Fukumura D, Jain RK, Seed B. Mouse embryonic fibroblasts exhibit extensive developmental and phenotypic diversity. *Proc Natl Acad Sci U S A*. 2016 Jan 5;113(1):122-7.
158. Paddison PJ, Caudy AA, Bernstein E, Hannon GJ, Conklin DS. Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. *Genes Dev*. 2002 Apr 15;16(8):948-58.
159. Kim DE, Procopio MG, Ghosh S, Jo SH, Goruppi S, Magliozzi F, Bordignon P, Neel V, Angelino P, Dotto GP. Convergent roles of ATF3 and CSL in chromatin control of cancer-associated fibroblast activation. *J Exp Med*. 2017 Aug 7;214(8):2349-2368.
160. Toyama EQ, Herzig S, Courchet J, Lewis TL Jr, Losón OC, Hellberg K, Young NP, Chen H, Polleux F, Chan DC, Shaw RJ. Metabolism. AMP-activated protein kinase mediates mitochondrial fission in response to energy stress. *Science*. 2016 Jan 15;351(6270):275-281.
161. Murray CW, Brady JJ, Tsai MK, Li C, Winters IP, Tang R, Andrejka L, Ma RK, Kunder CA, Chu P, Winslow MM. An LKB1-SIK Axis Suppresses Lung Tumor Growth and Controls Differentiation. *Cancer Discov*. 2019 Nov;9(11):1590-1605.
162. Alexander A, Cai SL, Kim J, Nanez A, Sahin M, MacLean KH, Inoki K, Guan KL, Shen J, Person MD, Kusewitt D, Mills GB, Kastan MB, Walker CL. ATM signals to TSC2 in the cytoplasm to regulate mTORC1 in response to ROS. *Proc Natl Acad Sci U S A*. 2010 Mar 2;107(9):4153-8.
163. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013 Jan 1;29(1):15-21.

164. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015 Jan 15;31(2):166-9.
165. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015 Apr 20;43(7):e47.
166. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010 Jan 1;26(1):139-40.
167. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005 Oct 25;102(43):15545-50.
168. Smedley D, Haider S, Durinck S, Pandini L, Provero P, Allen J, Arnaiz O, Awedh MH, Baldock R, Barbiera G, Bardou P, Beck T, Blake A, Bonierbale M, Brookes AJ, Bucci G, Buetti I, Burge S, Cabau C, Carlson JW, Chelala C, Chrysostomou C, Cittaro D, Collin O, Cordova R, Cutts RJ, Dassi E, Di Genova A, Djari A, Esposito A, Estrella H, Eyras E, Fernandez-Banet J, Forbes S, Free RC, Fujisawa T, Gadaleta E, Garcia-Manteiga JM, Goodstein D, Gray K, Guerra-Assunção JA, Haggarty B, Han DJ, Han BW, Harris T, Harshbarger J, Hastings RK, Hayes RD, Hoede C, Hu S, Hu ZL, Hutchins L, Kan Z, Kawaji H, Keliet A, Kerhornou A, Kim S, Kinsella R, Klopp C, Kong L, Lawson D, Lazarevic D, Lee JH, Letellier T, Li CY, Lio P, Liu CJ, Luo J, Maass A, Mariette J, Maurel T, Merella S, Mohamed AM, Moreews F, Nabihoudine I, Ndegwa N, Noirot C, Perez-Llamas C, Primig M, Quattrone A, Quesneville H, Rambaldi D, Reecy J, Riba M, Rosanoff S, Saddiq AA, Salas E, Sallou O, Shepherd R, Simon R, Sperling L, Spooner W, Staines DM, Steinbach D, Stone K, Stupka E, Teague JW, Dayem Ullah AZ, Wang J, Ware D, Wong Erasmus M, Youens-Clark K, Zadissa A, Zhang SJ, Kasprzyk A. The BioMart community portal: an innovative alternative to large, centralized data repositories. Nucleic Acids Res. 2015 Jul 1;43(W1):W589-98.
169. Gonzalez-Hunt CP, Rooney JP, Ryde IT, Anbalagan C, Joglekar R, Meyer JN. PCR-Based Analysis of Mitochondrial DNA Copy Number, Mitochondrial DNA Damage, and Nuclear DNA Damage. *Curr Protoc Toxicol*. 2016 Feb 1;67:20.11.1-20.11.25.
170. Contreras CM, Akbay EA, Gallardo TD, Haynie JM, Sharma S, Tagao O, Bardeesy N, Takahashi M, Settleman J, Wong KK, Castrillon DH. Lkb1 inactivation is sufficient to drive endometrial cancers that are aggressive yet highly responsive to mTOR inhibitor monotherapy. *Dis Model Mech*. 2010 Mar-Apr;3(3-4):181-93.

11. Appendixes

GEO Identifier	Direct/ Indirect	Comparison	DIrectPatientDerived **CAF**	IndirectPatientDerived **CAF**	Material	Organ	Cell type	Organism	Diesea model	Variables	Log 2FC	**Adj-P-Value**
GSE1724	Indirect	FASSc-vs-CNTRL		+	Patient derived cell line	Lung	Fibroblast	Human	Idiopathic pulmonary fibrosis and fibrosing alveolitis associated with systemic sclerosis	NA	-1.11E-01	**0.993**
GSE1724	Indirect	IPF-vs-CNTRL		+	Patient derived cell line	Lung	Fibroblast	Human	Idiopathic pulmonary fibrosis and fibrosing alveolitis associated with systemic sclerosis	NA	-0.362823	**0.905**
GSE17978	**Indirect**	**IPF-vs-CNTRL**	**+**		**Patient derived cell line**	**Lung**	**Fibroblast**	**Human**	**Idiopathic pulmonary disease**	**Overall - RNA extraction after 45 minutes of plating -> DIRECT**	-0.474213	**0.0285**
GSE22862	Indirect	CAF-vs-CNTRL		+	Patient derived cell line	Lung	Fibroblast	Human	NSCLC	NA	-0.079898	**0.925774**
GSE22863	**Direct**	**CAF-vs-CNTRL**	**+**		**Laser capture microdissected**	**Lung**	**Fibroblast**	**Human**	**NSCLC**	**NA**	0.132	**0.281**
GSE40839	Indirect	IntpneumCAF-vs-CNTRL		+	Patient derived cell line	Lung	Fibroblast	Human	Interstitial lung disease associated with systemic sclerosis and interstitial pneumonia	NA	-0.250998	**0.601528**
GSE40839	Indirect	ScleroCAF-vs-CNTRL		+							0.135441	**0.570362**
GSE44723	Indirect	rapidIPF-vs-nonIPF		+	Patient derived cell line	Lung	Fibroblast	Human	Idiopathic pulmonary disease	NA	0.097969	**0.64742**
GSE44723	Indirect	slowIPF-vs-nonIPF		+	Patient derived cell line	Lung	Fibroblast	Human	Idiopathic pulmonary disease	NA	0.037726	**0.95752**

GSE45686	Indirect	IPF-vs-CNTRL		+	Patient derived cell line	Lung	Fibroblast	Human	Idiopathic pulmonary disease	IPF grown in **control ECM** versus control fibroblast grown in **control ECN**	-0.236	**0.999**
GSE71351	Indirect	IPF_SLOW-vs-CNTRL		+	Patient derived cell line	Lung	Fibroblast	Human	Idiopathic pulmonary disease	Slow	0.0789	**0.997**
GSE71351	Indirect	IPF_FAST-vs-CNTRL		+	Patient derived cell line	Lung	Fibroblast	Human	Idiopathic pulmonary disease	Fast	-0.330131	**0.752**
GSE79786	Indirect	PAH_PERICYTES-vs-CNTRL		+	Patient derived cell line	Lung	Pericyte	Human	Pulmonary aterial hypertension	NA	0.373	**0.905**
GSE116679	Indirect	HIGH-vs-LOW		+	Patient derived cell line	Lung	CAF	Human	NSCLC	High vs low desmoplasia	-0.131248	**0.997**
GSE48397	Indirect	CAF-vs-CNTRL		+	Patient derived cell line	Lung	Fibroblast	Mouse	NSCLC	NA	-0.070812	**0.500325**
GSE8977	**Direct**	**IDC-vs-DSIC (INVS-vs-nonINVS)**	**+**		**Patient sample**	**Breast**	**Stroma**	**Human**	**Breast cancer**	**NA**	-0.113479	**0.680303**
GSE9014	**Direct**	**CAF-vs-CNTRL**	**+**		**Laser capture microdissected**	**Breast**	**Stroma**	**Human**	**Breast cancer**	**Invasive Ductal Carcinoma (IDC) samples; mixed mutations**	-0.789999	**1.29E-05**
GSE14548	**Direct**	**DCIS-vs-CNTRL**	**+**		**Laser capture microdissected**	**Breast**	**Stroma**	**Human**	**Breast cancer**	**NA**	-0.167	**0.466**
GSE14548	**Direct**	**IDC-vs-CNTRL**	**+**		**Laser capture microdissected**	**Breast**	**Stroma**	**Human**	**Breast cancer**	**NA**	0.090474	**0.907**
GSE20086	Indirect	CAF-vs-CNTRL		+	Patient derived cell line	Breast	Flbroblast	Human	Breast cancer	NA	-0.234675	**0.300981**
GSE26910	**Direct**	**CAF-vs-CNTRL**	**+**		**Microdissected**	**Breast**	**Stroma**	**Human**	**Breast cancer**	**NA**	-0.298495	**0.573132**

GSE 26910	**Direct**	**CAF-vs-CNTRL**	**+**		**Microdissected**	**Prostate**	**Stroma**	**Human**	**Prostate cancer**	**NA**	**-0.13221**	**0.984505**
GSE 29270	Indirect	CAF-vs-CNTRL		+	Patient derived cell line	Breast	Fibroblast	Human	Breast cancer	NA	-0.125543	0.19428
GSE 45256	Indirect	HYPERPLASIA-vs-CNTRL		+	Patient derived cell line	Breast	Fibroblast	Mouse	Breast cancer	HYPERPLASIA: MMTV-PyMT murine breast cancer model	-0.0667	**0.912209**
GSE 45256	Indirect	ADENOMA-vs-CNTRL		+	Patient derived cell line	Breast	Fibroblast	Mouse	Breast cancer	ADENOMA: MMTV-PyMT murine breast cancer model	0.09124 7	**0.92677**
GSE 45256	Indirect	CARCINOMA-vs-CNTRL		+	Patient derived cell line	Breast	Fibroblast	Mouse	Breast cancer	CARCINOMA: MMTV-PyMT murine breast cancer model	0.0282	**0.98799**
GSE 129189	**Direct**	**HYPERPLASIA-vs-CNTRL**	**+**		**FACS sorted fibroblasts**	**Breast**	**Fibroblast**	**Mouse**	**Breast cancer**	**FVB/N-MMTV-PyMT murine breast cancer model**	0.14522	**0.870174**
GSE 129189	**Direct**	**CARCINOMA-vs-CNTRL**	**+**		**FACS sorted fibroblasts**	**Breast**	**Fibroblast**	**Mouse**	**Breast cancer**	**FVB/N-MMTV-PyMT murine breast cancer model**	0.375184	**0.269441**
GSE 129189	**Direct**	**LateCARCINOMA-vs-CNTRL**	**+**		**FACS sorted fibroblasts**	**Breast**	**Fibroblast**	**Mouse**	**Breast cancer**	**FVB/N-MMTV-PyMT murine breast cancer model**	0.36588 1	**0.246062**
GSE 114056	Indirect	CAF-vs-CNTRL		+	Patient derived cell line	Breast	Fibroblast	Mouse	Breast cancer	MMTV-PyMT, DKK3 RNAi	0.29868 1	**0.562385**
GSE 5847	**Direct**	**IBC-vs-nonIBC**	**+**		**Laser capture microdissected**	**Breast**	**CAF**	**Human**	**Inflammatory Breast Cancer**	**IBC-vs-nonIBC**	0.23876 2	**0.25603**
GSE 10797	**Direct**	**CAF-vs-CNTRL**	**+**		**Laser capture**	**Breast**	**Stroma**	**Human**	**Breast cancer**	**NA**	0.05956 3	**0.967298**

					microdissected							
GSE83591	**Direct**	**CAF-vs-CNTRL**	**+**		**Laser capture microdissected**	**Breast**	**CAF**	**Huma**	**Breast cancer**	**HER+ breast cancer**	0.54963 7	**0.0541**
GSE90505	**Direct**	**CAF-vs-CNTRL**	**+**		**Laser capture microdissected**	**Breast**	**CAF**	**Huma**	**Breast cancer**	**Triple negative**	0.52080 5	**0.105**
GSE38666	**Direct**	**CAF-vs-CNTRL**	**+**		**Laser capture microdissected**	**Ovary**	**Stroma**	**Human**	**Ovarian cancer**	**NA**	0.23311 1	**0.45194**
GSE40595	**Direct**	**CAF-vs-CNTRL**	**+**		**Laser capture microdissected**	**Ovary**	**Stroma**	**Human**	**Ovarian cancer**	**NA**	-0.87572	**0.000527**
GSE115635	**Direct**	**CAF-vs-CNTRL**	**+**		**Laser capture microdissected**	**Ovary**	**Stroma**	**Human**	**Ovarian cancer**	**NA**	-0.59736 9	**0.044265**
GSE118624	Indirect	CA_MSC-vs-OM_MSC		+	Patient derived cell line	Ovary and omentum	CA-MSC	Human	Ovarian cancer	CA-MSC (ovary) vs MSC (omentum)	-0.18940 5	**0.219229**
GSE29156	**Direct**	**ADJC-vs-DIST_BENIGN**	**+**		**Laser capture microdissected**	**Cervix**	**Stroma**	**Human**	**Cervical cancer**	**NA**	-0.303	**0.99781**
GSE29156	**Direct**	**ADJC-vs-DIST_MALIGN**	**+**		**Laser capture microdissected**	**Cervix**	**Stroma**	**Human**	**Cervical cancer**	**NA**	0.35952 3	**0.999882**
GSE102232	**Direct**	**CAF-vs-CNTRL**	**+**		**Laser capture microdissected**	**Cervix**	**Stroma**	**Mouse**	**Cervical cancer**	**CAF = MUT_STROMA_EST - MUT_STROMA <= ESTROGEN EFFECT IN MUTANT MOUSE (HERE**	-0.12909	**0.722863**

										ESTROGEN CAUSES CANCER!)		
GSE35364	Indirect	CAF-vs-CNTRL		+	Patient derived cell line	Omentum	Fibroblast	Human	Ovarian cancer	Omental metastasis CAF vs normal adjecant fibroblast	-0.24362	**0.359168**
GSE35364	Indirect	coCAF-vs-CNTRL		+	Patient derived cell line	Omentum	Fibroblast	Human	Ovarian cancer	Normal omental fibroblast coculture d with cancer (ovarian) cells vs normal adjecant fibroblast	0.41758	**0.001133**
GSE31279	**Direct**	**CAF-vs-CNTRL**	**+**		**Laser capture microdissected**	**Colon**	**Stroma**	**Human**	**Colorectal cancer**		-0.185	**0.74388**
GSE35602	**Direct**	**CAF-vs-CNTRL**	**+**		**Laser capture microdissected**	**Colon**	**Stroma**	**Human**	**Colorectal cancer**		1.518891	**0.000875**
GSE46824	Indirect	CAF-vs-CNTRL		+	Patient derived cell line	Colon	Fibroblast	Human	Colorectal cancer		-0.205595	**0.264281**
GSE67250	Indirect	CHRON-vs-CNTRL		+	Patient derived cell line	Colon	Fibroblast	Human	Chron's disease		-0.132476	**0.816279**
GSE70468	Indirect	CAF-vs-CNTRL		+	Patient derived cell line	Colon	Fibroblast	Human	Colorectal cancer	Vitamin D (1,25(OH)2D3)	0.08961	**0.914**
GSE83834	Indirect	CAF-vs-CNTRL		+	Patient derived cell line	Stomach	Fibroblast	Human	Diffuse type gastric cancer	NA	-0.048013	**0.733959**
GSE90607	Indirect	stenoCAF-vs-CNTRL		+	Patient derived cell line	Ileum	Fibroblast	Human	Chron's disease	NA	8.27E-05	**1**
GSE90607	Indirect	inflCAF-vs-CNTRL		+							0.102	**1**
GSE99816	Indirect	CAF_NONINFL-vs-CNTRL		+	Patient derived cell line	Ileum	Fibroblast	Human	Chron's disease	Non-inflamed	-0.145076	**0.996782**
GSE99816	Indirect	CAF_INFL-vs-CNTRL		+	Patient derived cell line	Ileum	Fibroblast	Human	Chron's disease	Inflamed	0.093699	**0.98259**

GSE99816	Indirect	CAF_STENO-vs-CNTRL		+	Patient derived cell line	Ileum	Fibroblast	Human	Chron's disease	Stenotic	0.01291	**0.976216**
GSE62740	Indirect	CAF-vs-CNTRL		+	Patient derived cell line	Stomach	Fibroblast	Human	Colorectal cancer	NA	-0.04783	**0.85504**
GSE116167	Indirect	HSC_44PE-vs-NF_CoCult		+	Patient derived cell line	Stomach	Fibroblast	Human	Diffuse type gastric cancer	Low-metastatic (HSC-44PE) Co-culture	0.19619	**0.57566**
GSE116167	Indirect	44As3-vs-NF_CoCult		+	Patient derived cell line	Stomach	Fibroblast	Human	Diffuse type gastric cancer	High-metastatic (44As3) Co-culture	0.07159	**0.913024**
GSE11682	**Direct**	**CAF-vs-CNTRL**	**+**		**Laser capture microdissected**	**Prostate**	**Stroma**	**Human**	**Prostate cancer**	**NA**	-0.77738	**1**
GSE17906	**Direct**	**CAF-vs-CNTRL**	**+**		**FACS sorted CAFs**	**Prostate**	**CAF**	**Human**	**Prostate cancer**	**Prostate**	1.05732	**0.129124**
GSE17906	**Direct**	**CAF-vs-CNTRL**	**+**		**FACS sorted CAFs**	**Bladder**	**CAF**	**Human**	**Bladder cancer**	**Bladder**	-0.37504	**0.439183**
GSE34312	Indirect	CAF-vs-CNTRL		+	Patient derived cell line	Prostate	Fibroblast	Human	Prostate cancer	NA	0.284	**0.751**
GSE85606	Indirect	CAF-vs-CNTRL		+	Patient derived cell line	Prostate	Fibroblast	Human	Prostate cancer	CAF-vs-CNTRL	0.00120	**0.999924**
GSE86256	Indirect	CAF-vs-CNTRL		+	Patient derived cell line	Prostate	Fibroblast	Human	Prostate cancer	CAF-vs-CNTRL	-0.02783	**1**
GSE68164	Indirect	CAF-vs-CNTRL		+	Patient derived cell line	Prostate	Fibroblast	Human	Prostate cancer	CAF-vs-CNTRL	0.11780	**0.999**
GSE41344	**Direct**	**JOCK_CID-vs-CNTRL**	**+**		**Laser capture microdissected**	**Prostate**	**CAF**	**Mouse**	**Prostate cancer**	**JOCK mouse**	0.72155	**0.4135**
GSE43770	Indirect	CAF-vs-CNTRL		+	Patient derived cell line	Pancreas	Pancreatic stellate cell	Human and mouse	Pancreatic cancer	Patient derived CAF versus CNTRL	0.58359	**0.300471**
GSE93313	indirect	TRANSWELL-vs-3D		+	Patient derived cell line	Pancreas	Pancreatic stellate cell	Mouse	Pancreatic cancer	KrasLSL-G12D; Trp53LSL-R162H; Pdx1-Cre derived pancreati	-0.14629	**0.421233**

										c stellate cells 3D-transwell system with pancreatic cancer cells: activation through cancer cells		
GSE101665	Indirect	CAF-vs-CNTRL		+	Patient derived cell line	Pancreas	CAF	Human	Pancreatic cancer	CAF-vs-CNTRL	-0.23240 1	**0.00204 7**
GSE106901	**Direct**	**CAF-vs-CNTRL**	**+**		**FACS sorted fibroblasts**	**Pancreas**	**Fibroblast**	**Mouse**	**Pancreatic cancer**	**CAF-vs-CNTRL : K-Ras+/+;Elas-tTA/tetO-Cre; Rosa26+/EYFP**	-1.32568 5	**0.14156 6**
GSE17817	**Direct**	**CAF-vs-CNTRL**	**+**		**FACS sorted fibroblasts**	**Eardermis**	**Fibroblast**	**Mouse**	**Skin dysplasia**	**NA**	-0.18414 4	**0.96998 8**
GSE37738	Indirect	cSCC-vs-CNTRL		+	Patient derived cell line	Dermis	Fibroblast	Human	Cutaneous squamous cell carcinoma	Cutaneous squamous cell carcinoma (cSCC)	0.42756	**1**
GSE37738	Indirect	RDEB-vs-CNTRL		+	Patient derived cell line	Dermis	Fibroblast	Human	Recessive dystrophic epidermolysis bullosa	Recessive dystrophic epidermolysis bullosa (RDEB)	0.37044	**1**
GSE37738	Indirect	cSCC_RDEB-vs-CNTRL		+	Patient derived cell line	Dermis	Fibroblast	Human	Both	cSCC and RDEB	0.19559 8	**1**
GSE71078	Indirect	CAF-vs-CNTRL		+	Patient derived cell line	Dermis	Fibroblast	Human	BRCA1 mosaic epimutation	NA	-0.00604 9	0.99556 9
GSE83314	Indirect	CAF-vs-CNTRL		+	Patient derived cell line	Dermis	Fibroblast	Human	Head and neck SCC	NA	0.13124 8	**1**
GSE122370	Indirect	CAF-vs-CNTRL		+	Patient derived cell line	Dermis	Fibroblast	Human	Dermal SCC	NA		
GSE38517	indirect	CAF_ML-vs-CNTRL		+	Patient derived cell line	Lingual mucosa	Fibroblast	Human	Oral SCC	CAFs grown in 2D versus NFs grown in 3D	-0.42491 6	**0.06854 4**
GSE38517	indirect	CAF-vs-CNTRL		+	Patient derived cell line	Lingual	Fibroblast	Human	Oral SCC	CAFs versus NFs	-0.15	**0.39402 3**

						mucosa					9384	
GSE38517	indirect	DYSPL-vs-CNTRL		+	Patient derived cell line	Lingual mucosa	Fibroblast	Human	Oral SCC	Dysplastic fibroblasts grown in 3D versus NFs grown in 3D	-0.151782	**0.465211**
GSE73728	Direct	day4_TDLN-vs-CNTRL		+	Patient derived cell line	Lymph node	Fibroblast reticular cell	Mouse	-	Tumor draining effects on lymph node. Compared to PBS draining. Day 4		
GSE73728	Direct	day11_TDLN-vs-CNTRL		+	Patient derived cell line	Lymph node	Fibroblast reticular cell	Mouse	-	Tumor draining effects on lymph node. Compared to PBS draining. Day 11		

Appendix I: Table of Published CAF-data sets. Table of published direct and indirect cancer stroma datasets used in figures Fig1A; Fig1B; Fig3 and FigS5. Color-coding according to tissue.

Gene_primer name orientation	Primer sequence 5'-3'
Mm_Oaz1b_F	**CGCACCATGCCGCTTCTTA**
Mm_Oaz1b_R	**ATCCCGCTGACTGTTCCCT**
Mm_Prkaa1_F	GCA AAA TGG CAT TTG GGA AT
Mm_Prkaa1_R	CGA TAG TTG CTG ATG GAT CC
Mm_Prkaa2_F	CAC AGG CAT ATG GTT GTC CA
Mm_Prkaa2_R	CGG ATC TTC TTG AAG AGC GT
Mm_Sik1D_F	TCATGTCGGAGTTCAGTGCG
Mm_Sik1D_R	ACCTGCGTTTTGGTGACTCG
Mm_Sik2_F	GGCCGGAAGGTAGTGCACAG
Mm_Sik2_R	GAATCTTCCTTCTAAAACCC
Mm_Sik3D_F	GCCATCCACACATCATCAGAC
Mm_Sik3D_R	CCAAGTGGTCAAATATCTCCCC
Mm_Mark1B_F	GAAGAAACTGTTAGTGCTGA
Mm_Mark1B_R	AGAAGGATGTAAGTAGCCAT
Mm_Mark2B_F	CTACCCACGCTGAACGAAAGG
Mm_Mark2B_R	GTAGTTGCCAATATGGGGCTG
Mm_Mark3B_F	CCCTTTGCCAACGGTGAATG
Mm_Mark3B_R	CCCTTGCCGATTGTTTTCAAC
Mm_Mark4B_F	GGACACGCATGGCACATTG
Mm_Mark4B_R	GCAGGAAGCGATAGAGTTCCG
Mm_Nuak1C_F	TCCAACCTGTACCAGAAGGAC
Mm_Nuak1C_R	GGGCATCGTTCCATAAATGAGA
Mm_Nuak2C_F	ATCAAGTCGCCTAAACCTCTGA
Mm_Nuak2C_R	AATCTCCCTCCGTATGTGCAG
Mm_SnrkB_F	TTTAGGCGAGGATATGATGGGA
Mm_SnrkB_R	GTCCAGCTTCGTCTTGTCAAT
Mm_Oaz1b_F	CGCACCATGCCGCTTCTTA
Mm_Oaz1b_R	ATCCCGCTGACTGTTCCCT
Mm_Lkb1_F	CCCTTTGGAGGGAAAAATTC
Mm_Lkb1_R	TTTATCCCTATTGCCCAGAAC
Hs_GAPDH_F	**CGACCACTTTGTCAAGCTCA**
Hs_GAPDH_R	**AGGGGAGATTCAGTGTGGTG**
Hs_LKB1_F	GCTCTTACGGCAAGGTGAAG
Hs_LKB1_R	TTTTGTGCCGTAACCTCCTC
Hs_VDR_exon4_1_F	GTGGACATCGGCATGATGAAG
Hs_VDR_exon5_1_R	GGTCGTAGGTCTTATGGTGGG
Hs_TGFBI_exon4_1_F	CTTCGCCCCTAGCAACGAG
Hs_TGFBI_exon5_1_R	TGAGGGTCATGCCGTGTTTC
Hs_ADAMTS5_exon1_1_F	ACTACGATGCAGCTATCCTGT
Hs_ADAMTS5_exon2_1_R	GTCCCAACGTCTGCCATTC
Hs_ADAMTS5_exon3_2_F	TCTAAGCCCTGGTCCAAATG
Hs_ADAMTS5_exon4_2_R	CAGATTCTCCCCTTTCCACA
Hs_LIMCH1_exon17_1_F	CAAAGCAGTGCCTATGCTGA
Hs_LIMCH1_exon18_1_R	GGCCACACCTTCAAACATCT
Hs_PTGS1_exon5_1_F	CTCTGTGCCTAAAGATTGCCC
Hs_PTGS1_exon7_1_R	GTCTCCATAAATGTGGCCGAG
Hs_CXCR4_exon1_1_F	ACTACACCGAGGAAATGGGCT

Hs_CXCR4_exon1_1_R	CCCACAATGCCAGTTAAGAAGA
Hs_ATP1A2_exon8_1_F	TTCTTCGTGCTCTCCCTCAT
Hs_ATP1A2_exon9_1_R	GGTTCTTCACCAGGCAGTTC
Hs_SNCA_exon2_1_F	AAGAGGGTGTTCTCTATGTAGGC
Hs_SNCA_exon4_1_R	GCTCCTCCAACATTTGTCACTT
Hs_ITM2A_exon3_1_F	AATGACTGCTTACCTGGACTTG
Hs_ITM2A_exon4_1_R	TCCACAGCAACTAGGTCTTCT
Hs_CDO1_exon3_1_F	ACATTATTTGCCTGGCCTGA
Hs_CDO1_exon4_1_R	CAAGTGAAGGCTCACAGCAG
Hs_HLA-MDA_exon2_1_F	GCCAAAACTTGATGGGAAAA
Hs_HLA-MDA_exon3_1_R	CCATCGACAGCTGAGACAAA
Hs_SLPI_exon2_1_F	GAAATCTGCCCAGTGCCTTA
Hs_SLPI_exon3_1_R	GCATCAAACATTGGCCATAA
Hs_C3_exon12_1_F	GACCCTCAACGTCAACTTCC
Hs_C3_exon13_1_R	GCACCGATCAGCGTGTAGTA
Hs_MAL2_exon2_1_F	TCTTTCTGGGCATGTTCCTC
Hs_MAL2_exon3_1_R	GCCCGGTTATGGTTGTATTG
Hs_KRT18_exon3_1_F	TCGCAAATACTGTGGACAATGC
Hs_KRT18_exon4_1_R	GCAGTCGTGTGATATTGGTGT
Hs_UCP2_exon5_1_F	GGAGGTGGTCGGAGATACCAA
Hs_UCP2_exon6_1_R	ACAATGGCATTACGAGCAACAT
Hs_CEMIP_exon13_1_F	TAGTGATGGGGGAGATGGAG
Hs_CEMIP_exon14_1_R	AGCTGCTGTCCCATATGCTT
Hs_THY1_exon3_1_F	GAACCAACTTCACCAGCAAA
Hs_THY1_exon4_1_R	GCCCTCACACTTGACCAGTT
Hs_IDH2_exon2_1_F	CCCGTATTATCTGGCAGTTCATC
Hs_IDH2_exon3_1_R	ATCAGTCTGGTCACGGTTTGG
Hs_STAT2_exon5_1_F	GAGCCAGCAACATGAGATTGA
Hs_STAT2_exon6_1_R	GCCTGGATCTTATATCGGAAGCA
Hs_MAP3K8_exon4_1_F	CTCCCCAAAATGGACGTTACC
Hs_MAP3K8_exon5_1_R	GGATTTCCACATCAGATGGCTTA
Hs_EMB_exon5_1_F	CATGGGAAAAACAAGCCATT
Hs_EMB_exon6_1_R	CCCCATCTTCCTCCAAAAGT
Hs_NLRC5_exon14_1_F	GCTGGAAGATGAAGGCTGTC
Hs_NLRC5_exon16_1_R	TGGGCAAACATGAAGATCAC
Hs_PLAU_exon5_1_F	GCTTGTCCAAGAGTGCATGGT
Hs_PLAU_exon6_1_R	CAGGGCTGGTTCTCGATGG
Hs_SLC7A7_exon2_1_F	CCTTTGGAGGATTCCTTGCTT
Hs_SLC7A7_exon3_1_R	GTTCCCCATTTGACATAGGCA
Hs_CDCA3_exon3_1_F	GGAGCAACTGGAGGGTCTTA
Hs_CDCA3_exon34_1_R	GCTCTGGGGGAAGATTTGAT
Hs_COL11A1_exon3_1_F	TGGTGTTGAGGTTGGGAGAT
Hs_COL11A1_exon4_1_R	TCAAGTGGTTTCGTGGTTTTC
Hs_SMAD9_exon5_1_F	GAGTTGGGGAGACATTCCAG
Hs_SMAD9_exon6_1_R	AGATGCTGCTGTCACTCACG
Hs_TGFBR3_exon10_1_F	CCAAGATGAATGGCACACAC
Hs_TGFBR3_exon11_1_R	ATTTCAGGTCGGGTGAACAG
Hs_IL7R_exon3_1_F	CCCTCGTGGAGGTAAAGTGC
Hs_IL7R_exon4_1_R	CCTTCCCGATAGACGACACTC

Hs_IL10RA_exon2_1_F	CCTCCGTCTGTGTGGTTTGAA
Hs_IL10RA_exon3_1_R	CACTGCGGTAAGGTCATAGGA
Hs_MUC1_exon2_1_F	CTCGGTCCCAGTCACCAG
Hs_MUC1_exon3_1_R	CCGAGGTGACATTGTGGA
Hs_LYZ_exon2_1_F	TCAATAGCCGCTACTGGTGTA
Hs_LYZ_exon3_1_R	ATCACGGACAACCCTCTTTGC
Hs_SPP1_exon5_1_F	GAAGTTTCGCAGACCTGACAT
Hs_SPP1_exon6_1_R	GTATGCACCATTCAACTCCTCG
Hs_MXRA5_exon3_1_F	CATCCCCGATGGAGCTTTAAG
Hs_MXRA5_exon4_1_R	TCAATGTGCAGCCTCATTAAGTT
Hs_CXCL14_exon2_1_F	CGCTACAGCGACGTGAAGAA
Hs_CXCL14_exon3_1_R	GTTCCAGGCGTTGTACCAC
Hs_SLC1A3_exon6_1_F	GTCATCGTCTTGCCACTCCT
Hs_SLC1A3_exon7_1_R	GTCCACGCCATTGTTCTCTT
Hs_SLC40A1_exon7_1_F	CCGGAGACAAGTCCTGAATC
Hs_SLC40A1_exon8_1_R	ACACCATTTATAATGCCTCTTTCA
Hs_COL1A1_exon2_1_F	GTGCGATGACGTGATCTGTGA
Hs_COL1A1_exon3_1_R	CGGTGGTTTCTTGGTCGGT
Hs_COL6A2_exon10_1_F	GAACGGGACCGATGGACAG
Hs_COL6A2_exon13_1_R	CCCTTGGCCCCGATTTCTC
Hs_IL1R1_exon6_1_F	GGCTGAAAAGCATAGAGGGAAC
Hs_IL1R1_exon7_1_R	CTGGGCTCACAATCACAGG
Hs_ORAI2_exon3_1_F	GTCACCTCTAACCACCACTCG
Hs_ORAI2_exon4_1_R	CGGGTACTGGTACTGCGTCT
Hs_SYK_exon3_1_F	CATGGAAAAATCTCTCGGGAAGA
Hs_SYK_exon4_1_R	GTCGATGCGATAGTGCAGCA
Hs_PKD2_exon6_1_F	TGGAAATTCGCATTCACAAA
Hs_PKD2_exon7_1_R	TGCTCAAAGTTGGGGAAAGT
Hs_BPTF_exon1_1_F	CTTCAGGAGCCATAGTACCTACA
Hs_BPTF_exon2_1_R	CAAGGGGCGGGATGTCTTTTT
Hs_SRSF7_exon2_1_F	CAGGATTTGCCTTTGTGGAA
Hs_SRSF7_exon3_1_R	GGGCAGGTGGTCTATCAAAA
Hs_SENP6_exon7_1_F	TCCTGTAAGGTTAAGTCGGCT
Hs_SENP6_exon8_1_R	AGATAGAGGAGGAGTAGGCTGAT
Hs_GOSR1_exon4_1_F	TACCAACAGTGCAGGTGTCC
Hs_GOSR1_exon5_1_R	TGATCCCATAAGATTCTCCCTTT
Hs_FZD7_exon1_1_F	GTGCCAACGGCCTGATGTA
Hs_FZD7_exon1_1_R	AGGTGAGAACGGTAAAGAGCG
Hs_PAPOLA_exon18_1_F	GCAGCATCTGTGACCAACAT
Hs_PAPOLA_exon19_1_R	CAACTCTGGAGACCGTAGGC
Hs_KPNB1_exon3_1_F	AGGCACAATATCAGCAGAGGT
Hs_KPNB1_exon4_1_R	ACTAGGCCGGTAAGTTTCTGTA
Hs_SNAI2_exon2_1_F	TGTGACAAGGAATATGTGAGCC
Hs_SNAI2_exon3_1_R	TGAGCCCTCAGATTTGACCTG
Hs_PWP1_exon3_1_F	GGATGACAGGACGCTTGATGA
Hs_PWP1_exon4_1_R	CGTAGACCGTAAGACCCAAGA
Hs_DKK3_exon2_1_F	AGGAACTGATGGAGGACACG
Hs_DKK3_exon3_1_R	TTCCAACCTTCGTGTCTGTG
Hs_RABL6_exon9_1_F	ATCCTCTGTGCCCCCTGTA
Hs_RABL6_exon10_R	CGGTCGTCAGGAACAAAGTC
Hs_CCL11_exon2_1_F	CCCCTTCAGCGACTAGAGAG

Hs_CCL11_exon3_1_R	TCTTGGGGTCGGCACAGAT
Hs_GFPT2_exon4_1_F	AGACACACTTCGGCATTGC
Hs_GFPT2_exon6_1_R	TTGGCGATGGTCTCTGTATCT
Hs_KCNE4_exon1_1_F	AGCAGCGGAGTTGTCAGAG
Hs_KCNE4_exon2_1_R	GCTGTTCAGAGGCTCCATTT
Hs_SERPINB9_exon2_1_F	AATGCAAGTGGTACTTTTGCCA
Hs_SERPINB9_exon3_1_R	AAGCCCGATGAATGTCTTCCT
Hs_TNFAIP6_exon2_1_F	TCTGGCAAATACAAGCTCACC
Hs_TNFAIP6_exon3_1_R	CTGCCCTTAGCCATCCATCC
Mm_RMB15_F	**GGACAGTTTTCTTGGGCAAC**
Mm_RMB15_F	**AGTTTGGCCCTGTGAGACAT**
Mm_12srRNA_F	AGGAGGCTGTTCTATAATCGATAAA
Mm_12srRNA_R	GATGGCGGTATATAGGCCGAA

Appendix II: PCR-primers. Blue: Murine Stk11 and ARKs (FigS8). Red: Human STK11 and GAPDH (Fig1C; Fig1D). Purple: Human ovarian cancer responsive genes (Fig1C). Brown: Murine mitochondrial DNA primers (Fig4C; FigS6E). Housekeeping primers written in bold.

Target	Identifier
shPRKAA1	TRCN0000220672
shPRKAA2	TRCN0000220719
shMARK1	TRCN0000024173
shMARK2	TRCN0000220661
shMARK3	TRCN0000221202
shMARK4	TRCN0000024281
shNUAK1	TRCN0000221208
shNUAK2	TRCN0000024271
shSIK1	TRCN0000221194
shSIK2	TRCN0000024288
shSIK3	TRCN0000079132
shSNRK	TRCN0000024290
shLKB1	TRCN0000221241
shLKB1_#8	TRCN0000000408
shLKB1_#9	TRCN0000000409
shLKB1_#11	TRCN0000000411

Appendix III: The RNAi consortium (TRC) construct identifiers for mouse (Yellow) and human (Green) LKB1 and ARKs.

```
# APPENDIX IV

#-----------------------------------------------------------#
# Mapping and read counting for mouse trascriptomic data #
# GRCm38.p6                                                 #
# Single-end reads                                          #
#-----------------------------------------------------------#

# Location of reference genomes:
ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/001/635/GCF_000001635.26_GRCm
38.p6/
#-> gunzip them
#--> samtools faidx *.fna
#--> mkdir genomedir_GRCm38.p6_RefSeq

module load biokit

STAR --runMode genomeGenerate --genomeDir
/wrk/pekkajpp/DONOTREMOVE/reference/GRCm38.p6_RefSeq/genomedir_GRCm38.p6_Ref
Seq --genomeFastaFiles
/wrk/pekkajpp/DONOTREMOVE/reference/GRCm38.p6_RefSeq/GCF_000001635.26_GRCm38
.p6_genomic.fna --sjdbGTFfile
/wrk/pekkajpp/DONOTREMOVE/reference/GRCm38.p6_RefSeq/GCF_000001635.26_GRCm38
.p6_genomic.gff --sjdbOverhang 100 --runThreadN 2

# Mapping
nohup STAR --genomeDir
/wrk/pekkajpp/DONOTREMOVE/reference/GRCm38.p6_RefSeq/genomedir_GRCm38.p6_Ref
Seq --readFilesIn
/wrk/pekkajpp/DONOTREMOVE/starjob2/filename_concatenated.fastq --
outFilterType BySJout --outFilterMultimapNmax 20 --alignSJoverhangMin 8 --
alignSJDBoverhangMin 1 --outFilterMismatchNmax 999 --
outFilterMismatchNoverLmax 0.04 --alignIntronMin 20 --alignIntronMax 1000000
--outFileNamePrefix filename_concatenated --chimSegmentMin 30 --outSAMtype
BAM Unsorted SortedByCoordinate --runThreadN 12 &

# Formatting
nohup samtools view -h -o filename_concatenated.sorted.sam
filename_concatenatedAligned.sortedByCoord.out.bam &

Read counting
#HTSeq
module load python-env/3.5.3
pip install htseq

nohup python -m HTSeq.scripts.count -s no -i gene -t exon
filename_concatenated.sorted.sam
/wrk/pekkajpp/DONOTREMOVE/reference/GRCm38.p6_RefSeq/GCF_000001635.26_GRCm38
.p6_genomic.gff > filename_counts.txt &

#-----------------------------------------------------------#
# Mapping and read counting for human trascriptomic data #
# GRCh38.p12                                                #
# Single-end reads                                          #
#-----------------------------------------------------------#

prefetch -v -O /wrk/pekkajpp/DONOTREMOVE/GSE_identifier SRR_identifier.sra
```

```
fastq-dump --outdir /wrk/pekkajpp/DONOTREMOVE/GSE_identifier --split-files
SRR_identifier.sra

# Check the file size
# https://stackoverflow.com/questions/25845057/unix-command-to-check-the-
filesize

# Count the read number
# https://www.biostars.org/p/139006/

echo $(cat yourfile.fastq|wc -l)/4|bc

# Mapping

nohup STAR --genomeDir
/wrk/pekkajpp/DONOTREMOVE/reference/GRCh38.p12_RefSeq/genomedir_GRCh38.p12_R
efSeq --readFilesIn
/wrk/pekkajpp/DONOTREMOVE/GSE_identifier/identifier_1.fastq --outFilterType
BySJout --outFilterMultimapNmax 20 --alignSJoverhangMin 8 --
alignSJDBoverhangMin 1 --outFilterMismatchNmax 999 --
outFilterMismatchNoverLmax 0.04 --alignIntronMin 20 --alignIntronMax 1000000
--outFileNamePrefix SRR_identifier --chimSegmentMin 30 --outSAMtype BAM
Unsorted SortedByCoordinate --runThreadN 12 &

# HTseq-count

nohup samtools view -h -o SRR_identifier.sorted.sam
SRR_identifierAligned.sortedByCoord.out.bam &

module load python-env/3.5.3
pip install htseq

nohup python -m HTSeq.scripts.count -s no -i gene -t exon
SRR_identifier.sorted.sam
/wrk/pekkajpp/DONOTREMOVE/reference/GRCh38.p12_RefSeq/GCF_000001405.38_GRCh3
8.p12_genomic.gff > SRR_identifier_counts.txt &

#-----------------------------------------------------------#
# Mapping and read counting for human trascriptomic data #
# GRCh38.p12                                                #
# Paired-end reads                                         #
#-----------------------------------------------------------#

Human, paired

prefetch -v -O /wrk/pekkajpp/DONOTREMOVE/GSE_identifier SRR_identifier.sra

fastq-dump --outdir /wrk/pekkajpp/DONOTREMOVE/GSE_identifier --split-files
SRR_identifier.sra

# Check the file size
# https://stackoverflow.com/questions/25845057/unix-command-to-check-the-
filesize

# Count the read number
# https://www.biostars.org/p/139006/

echo $(cat yourfile.fastq|wc -l)/4|bc
```

```
# Mapping

nohup STAR --genomeDir
/wrk/pekkajpp/DONOTREMOVE/reference/GRCh38.p12_RefSeq/genomedir_GRCh38.p12_R
efSeq --readFilesIn /wrk/pekkajpp/DONOTREMOVE/GSE76125/filename_1.fastq
filename_2.fastq --outFilterType BySJout --outFilterMultimapNmax 20 --
alignSJoverhangMin 8 --alignSJDBoverhangMin 1 --outFilterMismatchNmax 999 --
outFilterMismatchNoverLmax 0.06 --alignIntronMin 20 --alignIntronMax 1000000
--outFileNamePrefix SRR3_filename --chimSegmentMin 30 --outSAMtype BAM
Unsorted SortedByCoordinate --runThreadN 12 &


# HTseq-count

nohup samtools view -h -o SRR_filename.sorted.sam
SRR_filenameAligned.sortedByCoord.out.bam &

module load python-env/3.5.3
pip install htseq

nohup python -m HTSeq.scripts.count -s no -i gene -t exon -r pos
SRR_filename.sorted.sam
/wrk/pekkajpp/DONOTREMOVE/reference/GRCh38.p12_RefSeq/GCF_000001405.38_GRCh3
8.p12_genomic.gff > SRR_filename_counts.txt &
```

Appendix IV: Unix-script for genome mapping, read counting and SRA-toolkit fetching.

```
# APPENDIX V
# Upload the counts
counts1<-read.delim(file = "counts1.txt", header = FALSE, col.names =
c('Transcript_ID','counts1'))
counts2<-read.delim(file = "counts2.txt", header = FALSE, col.names =
c('Transcript_ID','counts2'))
counts3<-read.delim(file = "counts3.txt", header = FALSE, col.names =
c('Transcript_ID','counts3'))
counts4<-read.delim(file = "counts4.txt", header = FALSE, col.names =
c('Transcript_ID','counts4'))

merged1<-merge(counts1,counts2,all = TRUE)
merged1<-merge(merged1,counts3,all = TRUE)
merged1<-merge(merged1,counts4,all = TRUE)

write.table(x = merged1, file = 'countmatrix.txt', sep = '\t', row.names =
FALSE, quote = FALSE)

# Annotation: This step might drop out some (significantly expressed
genes!!!)
#------------------------------------------------------------------------
dataframe <- merged1
library(biomaRt)
ensembl = useMart("ensembl")
ensembl = useMart("ensembl")
datasets = listDatasets(ensembl)
ensembl = useDataset("mmusculus_gene_ensembl", useMart("ensembl"))
genes = dataframe$Transcript_ID
data = dataframe[,2:5]
G_list <- getBM(attributes=
c('mgi_symbol','entrezgene','ensembl_gene_id','ensembl_transcript_id','descr
iption'),
                filters= 'mgi_symbol',
                values=genes,
                mart= ensembl)
mergedData <- merge(x = dataframe,
                    y = G_list,
                    by.x = "Transcript_ID",
                    by.y = "mgi_symbol"
)
write.table(mergedData, "/Volumes//merged_annotated.txt", sep = "\t")

#------------------------------------------------------------------------
# Differential expression with edgeR/Limma

library(edgeR)
rawData2 <- merged1
rawData2 <- unique(rawData2)

# We then create a DGEList object:
y <- DGEList(counts = rawData2[,2:5], genes = rawData2[,1])
y$samples
# Filtering
keep <- rowSums(cpm(y)>0.2) >= 4
table(keep)
```

```
y <- y[keep, , keep.lib.sizes = FALSE]

# Normalization
y <- calcNormFactors(y)
y$samples

# PCA (edgeR)
plotMDS(y, col=rep(1:4, each=4))

# Design matrix
exp_design <- data.frame(row.names = colnames(rawData2[,2:5]),
                         names = colnames(rawData2[,2:5]),
                         kd = factor(c(rep("wt", 2), rep("mutant", 2)))
)
exp_design$kd

design <- model.matrix(~0+exp_design$kd)

rownames(design) <- colnames(y)
colnames(design) <- c("mutant", "wt")

# DE
vwts <- voomWithQualityWeights(y, design = design, normalize.method =
"quantile", plot = FALSE)
fit <- lmFit(vwts, design)

# Make contrasts:
my.contrasts <- makeContrasts(
  mutant = mutant - wt,
  levels = design
)

fit.de <- contrasts.fit(fit, my.contrasts)
fit.de <- eBayes(fit.de)

summary(decideTests(fit.de, adjust.method = "fdr", p.value = 0.05))

for (i in 1:1) {
  res <- topTable(fit.de, coef = i, number = Inf, adjust.method = "fdr")
  #require(xlsx)
  write.table(res, file = paste(colnames(my.contrasts)[i], "vsWT_genes.txt",
sep = ""), row.names = TRUE, sep = "\t", quote = FALSE)
}

data <- read.table(file = 'mutantvsWT_genes.txt', header = TRUE, sep = '\t',
stringsAsFactors = FALSE)

#-----------------------------------------------------------------------
# Ranked file for GSEA

df <- data
attach(df)
df <- df[order(genes),]
df <- df[c(1,2)]
df <- unique(df)
write.table(x = df, file = "df.musmusculus.rnk.txt", sep = "\t", quote =
FALSE, row.names = FALSE)
```

Appendix V: R-script for differential expression analysis.

```
#APPENDIX VI

library(readxl)
library(dplyr)

# Input files:
AMPKa1 <- read.delim("AMPKa1vsSCR_genes_20190323.txt", check.names = FALSE,
stringsAsFactors = FALSE)
AMPKa1 <- AMPKa1[,c(1,2,6)]
AMPKa1 <- data.frame(AMPKa1)
AMPKa2 <- read.delim("AMPKa2vsSCR_genes_20190323.txt", check.names = FALSE,
stringsAsFactors = FALSE)
AMPKa2 <- AMPKa2[,c(1,2,6)]
AMPKa2 <- data.frame(AMPKa2)
LKB1 <- read.table("LKB1vsSCR_genes_20190323.txt", check.names = FALSE,
stringsAsFactors = FALSE)
LKB1 <- LKB1[,c(1,3,7)]
LKB1 <- data.frame(LKB1)

# Copy the excel sheet to clipboard
All_genes_all_KD_ST_RNAseqdata_merged <- read_excel(path =
'All_genes_all_KD_ST_RNAseqdata_merged.xlsx')
All_genes_all_KD_ST_RNAseqdata_merged <-
data.frame(All_genes_all_KD_ST_RNAseqdata_merged)
# If you import with 'read_excel' you need to transform the data into
data.frame object.
ST_LKB1 <- All_genes_all_KD_ST_RNAseqdata_merged[, c(1, 5, 6)]
ST_LKB1 <- data.frame(ST_LKB1)
NUAK1 <- All_genes_all_KD_ST_RNAseqdata_merged[, c(1, 8, 9)]
NUAK2 <- All_genes_all_KD_ST_RNAseqdata_merged[, c(1, 11, 12)]
MRK1 <- All_genes_all_KD_ST_RNAseqdata_merged[, c(1, 14, 15)]
MRK2 <- All_genes_all_KD_ST_RNAseqdata_merged[, c(1, 17, 18)]
MRK3 <- All_genes_all_KD_ST_RNAseqdata_merged[, c(1, 20, 21)]
MRK4 <- All_genes_all_KD_ST_RNAseqdata_merged[, c(1, 23, 24)]
SIK1 <- All_genes_all_KD_ST_RNAseqdata_merged[, c(1, 26, 27)]
SIK2 <- All_genes_all_KD_ST_RNAseqdata_merged[, c(1, 29, 30)]
SIK3 <- All_genes_all_KD_ST_RNAseqdata_merged[, c(1, 32, 33)]
SNRK <- All_genes_all_KD_ST_RNAseqdata_merged[, c(1, 35, 36)]

##############################################################################
# These vectors include only genes unique for LKB1_1 and LKB1_2 dataframes
respectively
unique_for_PP <- setdiff(LKB1[,1], ST_LKB1[,1])
length(unique_for_PP)
unique_for_PP_frame <- LKB1[LKB1[,1] %in% unique_for_PP, ]
#2235
unique_for_ST <- setdiff(ST_LKB1[,1], LKB1[,1])
length(unique_for_ST)
unique_for_ST_frame <- ST_LKB1[ST_LKB1[,1] %in% unique_for_ST, ]
#1846
colnames(unique_for_ST_frame) <- colnames(unique_for_PP_frame)

# These FILTERED dataframes include only common genes
LKB1_filtered <- LKB1[!(LKB1[,1] %in% unique_for_PP),]
ST_LKB1_filtered <- ST_LKB1[!(ST_LKB1[,1] %in% unique_for_ST),]
# Total common de genes:
# 13139
colnames(ST_LKB1_filtered) <- colnames(LKB1_filtered)
```

```
# Sorting the dataframes according to gene names
LKB1_sorted <- with(LKB1_filtered, LKB1_filtered[order(LKB1_filtered[,1]) ,
])
ST_LKB1_sorted <- with(ST_LKB1_filtered,
ST_LKB1_filtered[order(ST_LKB1_filtered[,1]) , ])
colnames(ST_LKB1_sorted) <- colnames(LKB1_sorted)  # Matching the col.names

LKB1_sorted <- LKB1_sorted[!duplicated(LKB1_sorted[,1]),]
rownames(LKB1_sorted) <- LKB1_sorted[,1]
LKB1_sorted <- LKB1_sorted[,c(2,3)]
ST_LKB1_sorted <- ST_LKB1_sorted[!duplicated(ST_LKB1_sorted[,1]),]
rownames(ST_LKB1_sorted) <- ST_LKB1_sorted[,1]
ST_LKB1_sorted <- ST_LKB1_sorted[,c(2,3)]

# Check for similarity between the two LKB1 datasets with respect to brute-
directionality
interceptUP <- NULL    # This 'll contain rows from highest upregalated LKB1
datasets REGARDLESS of p-value.
interceptDOWN <- NULL  # This 'll contain rows from lowest downregulated
LKB1 datasets REGARDLESS of p-value.
for (i in 1:(dim(LKB1_sorted)[1])) {
    if (((ST_LKB1_sorted[i,1] >= 0) & (LKB1_sorted[i,1] >= 0)) &
(ST_LKB1_sorted[i,1] >= LKB1_sorted[i,1])) {
      interceptUP <- rbind(interceptUP, ST_LKB1_sorted[i,])
    } else if (((ST_LKB1_sorted[i,1] >= 0) & (LKB1_sorted[i,1] >= 0)) &
(ST_LKB1_sorted[i,1] < LKB1_sorted[i,1])) {
      interceptUP <- rbind(interceptUP, LKB1_sorted[i,])
    } else if (((ST_LKB1_sorted[i,1] <= 0) & (LKB1_sorted[i,1] <= 0)) &
(ST_LKB1_sorted[i,1] > LKB1_sorted[i,1])) {
      interceptDOWN <- rbind(interceptDOWN, LKB1_sorted[i,])
    } else if (((ST_LKB1_sorted[i,1] <= 0) & (LKB1_sorted[i,1] <= 0)) &
(ST_LKB1_sorted[i,1] <= LKB1_sorted[i,1])) {
      interceptDOWN <- rbind(interceptDOWN, ST_LKB1_sorted[i,])
    }
}
common <- rbind(interceptUP, interceptDOWN)
# interceptUP: 4716
# interceptDOWN:  4645
# Total: 9361 -> 71.24591% show the same directionality

#############################################################################
# COMPARING HOW DOES ALL THE SUBSTRATES CONTRIBUTE TO THE OVERALL LKB1-
SIGNATURE TOGETHER AND INDIVIDUALLY (UNIQUELY EXPRESSED GENES WITHIN SINGLE
SUBSTRATE)

LKB1_adjPVal0.05_logFC0.59 <- subset(LKB1, ((LKB1[,2] >= 0.59 & LKB1[,3] <=
0.05) | (LKB1[,2] <= -1 & LKB1[,3] <= 0.05)))
ST_LKB1_adjPVal0.05_logFC0.59 <- subset(ST_LKB1, ((ST_LKB1[,2] >= 0.59 &
ST_LKB1[,3] <= 0.05) | (ST_LKB1[,2] <= -1 & ST_LKB1[,3] <= 0.05)))

signfUniqPekka <- setdiff(LKB1_adjPVal0.05_logFC0.59[,1],
ST_LKB1_adjPVal0.05_logFC0.59[,1])
signfUniqPekka <- LKB1[LKB1[,1] %in% signfUniqPekka, ]
signfUniqPekka <- signfUniqPekka[order(signfUniqPekka[,1]) , ]

signfUniqSushil <- setdiff(ST_LKB1_adjPVal0.05_logFC0.59[,1],
LKB1_adjPVal0.05_logFC0.59[,1])
```

```
signfUniqSushil <- ST_LKB1[ST_LKB1[,1] %in% signfUniqSushil, ]
signfUniqSushil <- signfUniqSushil[order(signfUniqSushil[,1]) , ]

signfCommonPekka <-
LKB1_adjPVal0.05_logFC0.59[!(LKB1_adjPVal0.05_logFC0.59[,1] %in%
signfUniqPekka[,1]) , ]
signfCommonPekka <- signfCommonPekka[order(signfCommonPekka[,1]) , ]
signfCommonSushil <-
ST_LKB1_adjPVal0.05_logFC0.59[!(ST_LKB1_adjPVal0.05_logFC0.59[,1] %in%
signfUniqSushil[,1]) , ]
signfCommonSushil <- signfCommonSushil[order(signfCommonSushil[,1]) , ]
colnames(signfCommonSushil) <- colnames(signfCommonPekka)

# How many of the "common" signicant genes are actually COMMON? That is,
that also share the same directioanlity!
commonSignfDOWN <- NULL  # 130 ## 25
commonSignfUP <- NULL  # 178 ## 178
for (i in 1:dim(signfCommonPekka)[1]) {
  for (j in 1:dim(signfCommonSushil)[1]) {
    if ((signfCommonPekka[i,1] == signfCommonSushil[j,1]) &
(signfCommonPekka[i,2] > 0) & (signfCommonSushil[j,2] > 0) &
(signfCommonPekka[i,2] > signfCommonSushil[j,2])) {
      commonSignfUP <- rbind(commonSignfUP, signfCommonPekka[i,])
    } else if ((signfCommonPekka[i,1] == signfCommonSushil[j,1]) &
(signfCommonPekka[i,2] > 0) & (signfCommonSushil[j,2] > 0) &
(signfCommonPekka[i,2] < signfCommonSushil[j,2])) {
      commonSignfUP <- rbind(commonSignfUP, signfCommonSushil[j,])
    } else if ((signfCommonPekka[i,1] == signfCommonSushil[j,1]) &
(signfCommonPekka[i,2] < 0) & (signfCommonSushil[j,2] < 0) &
(signfCommonPekka[i,2] < signfCommonSushil[j,2])) {
      commonSignfDOWN <- rbind(commonSignfDOWN, signfCommonPekka[i,])
    } else if ((signfCommonPekka[i,1] == signfCommonSushil[j,1]) &
(signfCommonPekka[i,2] < 0) & (signfCommonSushil[j,2] < 0) &
(signfCommonPekka[i,2] > signfCommonSushil[j,2])) {
      commonSignfDOWN <- rbind(commonSignfDOWN, signfCommonSushil[j,])
    }
  }
}

# nonSignfPekka <- LKB1[!(LKB1[,1] %in% LKB1_adjPVal0.05_logFC0.59[,1]), ]
# nonSignfPekka <- nonSignfPekka[order(nonSignfPekka[,1]) , ]
# nonSignfSushil <- ST_LKB1[!(ST_LKB1[,1] %in%
ST_LKB1_adjPVal0.05_logFC0.59[,1]), ]
# nonSignfSushil <- nonSignfSushil[order(nonSignfSushil[,1]) , ]
# colnames(nonSignfSushil) <- colnames(nonSignfPekka)

# How many of uniquely significant genes in Sushil's LKB1 dataset is
directionally similar to non-significant genes in  my LKB1 dataset
commonDOWN_Pekka <- NULL  # 40
commonUP_Pekka <- NULL    # 88
nonSignfPekka <- LKB1[(LKB1[,1] %in% signfUniqSushil[,1]) , ]
nonSignfPekka <- nonSignfPekka[order(nonSignfPekka[,1]) , ]
for (i in 1:dim(nonSignfPekka)[1]) {
  for (j in 1:dim(signfUniqSushil)[1]) {
    if ((nonSignfPekka[i,1] == signfUniqSushil[j,1]) & (nonSignfPekka[i,2]
>= 0) & (signfUniqSushil[j,2] > 0)) {
      commonUP_Pekka <- rbind(commonUP_Pekka, nonSignfPekka[i,])
```

```
    } else if ((nonSignfPekka[i,1] == signfUniqSushil[j,1]) &
(nonSignfPekka[i,2] <= 0) & (signfUniqSushil[j,2] < 0)) {
      commonDOWN_Pekka <- rbind(commonDOWN_Pekka, nonSignfPekka[i,])
    }
  }
}
# How many of uniquely significant genes in my LKB1 dataset is directionally
similar to non-significant genes in Sushil's LKB1 dataset
commonDOWN_Sushil <- NULL  # 91
commonUP_Sushil <- NULL  # 271
nonSignfSushil <- ST_LKB1[(ST_LKB1[,1] %in% signfUniqPekka[,1]) , ]
colnames(nonSignfSushil) <- colnames(nonSignfPekka)
for (i in 1:dim(nonSignfSushil)[1]) {
  for (j in 1:dim(signfUniqPekka)[1]) {
    if ((nonSignfSushil[i,1] == signfUniqPekka[j,1]) & (nonSignfSushil[i,2]
>= 0) & (signfUniqPekka[j,2] > 0)) {
      commonUP_Sushil <- rbind(commonUP_Sushil, nonSignfSushil[i,])
    } else if ((nonSignfSushil[i,1] == signfUniqPekka[j,1]) &
(nonSignfSushil[i,2] <= 0) & (signfUniqPekka[j,2] < 0)) {
      commonDOWN_Sushil <- rbind(commonDOWN_Sushil, nonSignfSushil[i,])
    }
  }
}

# This vector will contain all the significantly and similarly expressed
(pVal<0.05,
# log2FC either <-1 or +0.59<) between LKB1_1 and LKB1_2 dataset AND
# similarly expressed genes between unique significant genes from either
LKB1_1 or LKB1_2 dataset
# and the non-significant (outside adjPVal0.05_logFC0.59) from the other
dataset:
LKB1comp <- rbind(commonSignfDOWN,
           commonSignfUP,
           commonDOWN_Pekka,
           commonUP_Pekka,
           commonDOWN_Sushil,
           commonUP_Sushil)
# Then use the 'commonWithLkb1' -function from below
commonWithLkb1(NUAK1, LKB1comp, pValue = 0.05, log2FC = 0.59)
## Total common up/down genes: 174 / 36 ( 210 )
## Total up+down genes in NUAK1 : 1573
## Fraction of all NUAK1 genes: 0.1335029
commonWithLkb1(NUAK2, LKB1comp, pValue = 0.05, log2FC = 0.59)
## Total common up/down genes: 112 / 8 ( 120 )
## Total up+down genes in NUAK2 : 847
## Fraction of all NUAK2 genes: 0.1416765
commonWithLkb1(MRK1, LKB1comp, pValue = 0.05, log2FC = 0.59)
## Total common up/down genes: 33 / 3 ( 36 )
## Total up+down genes in MRK1 : 433
## Fraction of all MRK1 genes: 0.08314088
commonWithLkb1(MRK2, LKB1comp, pValue = 0.05, log2FC = 0.59)
## Total common up/down genes: 179 / 36 ( 215 )
## Total up+down genes in MRK2 : 2479
## Fraction of all MRK2 genes: 0.08672852
commonWithLkb1(MRK3, LKB1comp, pValue = 0.05, log2FC = 0.59)
## Total common up/down genes: 73 / 12 ( 85 )
## Total up+down genes in MRK3 : 1176
##Fraction of all MRK3 genes: 0.07227891
```

```
commonWithLkb1(MRK4, LKB1comp, pValue = 0.05, log2FC = 0.59)
## Total common up/down genes: 25 / 1 ( 26 )
## Total up+down genes in MRK4 : 406
## Fraction of all MRK4 genes: 0.06403941
commonWithLkb1(SIK1, LKB1comp, pValue = 0.05, log2FC = 0.59)
## Total common up/down genes: 71 / 17 ( 88 )
## Total up+down genes in SIK1 : 1261
## Fraction of all SIK1 genes: 0.06978588
commonWithLkb1(SIK2, LKB1comp, pValue = 0.05, log2FC = 0.59)
## Total common up/down genes: 31 / 4 ( 35 )
## Total up+down genes in SIK2 : 773
## Fraction of all SIK2 genes: 0.04527814
commonWithLkb1(SIK3, LKB1comp, pValue = 0.05, log2FC = 0.59)
## Total common up/down genes: 71 / 6 ( 77 )
## Total up+down genes in SIK3 : 946
## Fraction of all SIK3 genes: 0.08139535
commonWithLkb1(SNRK, LKB1comp, pValue = 0.05, log2FC = 0.59)
## Total common up/down genes: 63 / 7 ( 70 )
## Total up+down genes in SNRK : 1091
## Fraction of all SNRK genes: 0.06416132
commonWithLkb1(AMPKa1, LKB1comp, pValue = 0.05, log2FC = 0.59)
## Total common up/down genes: 113 / 14 ( 127 )
## Total up+down genes in AMPKa1 : 2006
## Fraction of all AMPKa1 genes: 0.06331007
commonWithLkb1(AMPKa2, LKB1comp, pValue = 0.05, log2FC = 0.59)
## Total common up/down genes: 157 / 22 ( 179 )
## Total up+down genes in AMPKa2 : 1150
## Fraction of all AMPKa2 genes: 0.1556522

x <- subset(SNRK, ((SNRK[,2] > 0.59 & SNRK[,3] <= 0.05) | (SNRK[,2] < -1 &
SNRK[,3] <= 0.05)))
write.table(x = x, file = 'SNRK.txt', quote = FALSE, sep = '\t', row.names =
FALSE, col.names = TRUE)

allGenesFromSubstrates <- c(AMPKa1commonAll, AMPKa2commonAll,
NUAK1commonAll, NUAK2commonAll, MRK1commonAll,
                            MRK2commonAll, MRK3commonAll, MRK4commonAll,
SIK1commonAll, SIK2commonAll, SIK3commonAll,
                            SNRKcommonAll)
allGenesFromSubstrates <- unique(allGenesFromSubstrates)
length(allGenesFromSubstrates)/dim(LKB1comp)[1]
# >>> 77.34% explained

#############################################################################
#This for checking the exclusive gene contribution of each substrate
# From this dataframe you can easily see what genes come from what
substrates!
df1 <- data.frame(col = NUAK1commonAll, NUAK1commonAll)
df2 <- data.frame(col = NUAK2commonAll, NUAK2commonAll)
df3 <- data.frame(col = MRK1commonAll, MRK1commonAll)
df4 <- data.frame(col = MRK2commonAll, MRK2commonAll)
df5 <- data.frame(col = MRK3commonAll, MRK3commonAll)
df6 <- data.frame(col = MRK4commonAll, MRK4commonAll)
df7 <- data.frame(col = SIK1commonAll, SIK1commonAll)
df8 <- data.frame(col = SIK2commonAll, SIK2commonAll)
df9 <- data.frame(col = SIK3commonAll, SIK3commonAll)
df10 <- data.frame(col = SNRKcommonAll, SNRKcommonAll)
df11 <- data.frame(col = AMPKa1commonAll, AMPKa1commonAll)
```

```
df12 <- data.frame(col = AMPKa2commonAll, AMPKa2commonAll)
allSubstrates <- Reduce(function(x, y) merge(x, y, all=TRUE), list(df1, df2,
df3, df4, df5, df6, df7, df8, df9, df10, df11, df12))
genes <- allSubstrates[,1]  # < This is just to set the 'allSubstrates' -
genes as the row names
genes <- as.character(genes) # < The same continues here...
row.names(allSubstrates) <- genes # < ...and here
allSubstrates <- allSubstrates[,c(2:13)] # Deleting the first column (all
genes)
allSubstrates <- data.matrix(allSubstrates) # data.frame >> data.matrix
allSubstrates <- ifelse(is.na(allSubstrates), 0, 1) # Renaming the values
from the data.matrix so that each hit is 1 and miss is 0.
rivisummat <- rowSums(allSubstrates) == 1 # Check which rows has only one
substrate!
table(rivisummat)
#rivisummat
#FALSE  TRUE
#513   268
rivisummat1 <- allSubstrates[rivisummat, ]
pylvassummat <- colSums(rivisummat1) # <<< This'll show how many UNIQUE
genes each substrate has together with LKB1

           # NUAK1commonAll  NUAK2commonAll   MRK1commonAll   MRK2commonAll
MRK3commonAll   MRK4commonAll   SIK1commonAll  SIK2commonAll   SIK3commonAll
SNRKcommonAll AMPKa1commonAll AMPKa2commonAll
# -1;+0.59 # 32              12               2              33
15              4               11             3              11
8             20              18

#############################################################################
# Function for substrate-wise contribution
commonWithLkb1 <- function(knockdown, comparedTo, pValue, log2FC) {
  name1 <- deparse(substitute(knockdown))

  downInKnock <- subset(knockdown, (knockdown[,2] < -(1) & knockdown[,3] <=
pValue))
  upInKnock <- subset(knockdown, (knockdown[,2] > log2FC & knockdown[,3] <=
pValue))
  downInLkb1Union <- subset(comparedTo, (comparedTo[,2] < 0))
  upInLkb1Union <- subset(comparedTo, (comparedTo[,2] > 0))
  commonDownInKnock <- downInKnock[(downInKnock[,1] %in%
downInLkb1Union[,1]),]
  commonUpInKnock <- upInKnock[(upInKnock[,1] %in% upInLkb1Union[,1]),]

  commonAllInKnock <- c(commonDownInKnock[,1], commonUpInKnock[,1])
  x <- (dim(commonDownInKnock)[1] +
dim(commonUpInKnock)[1])/(dim(downInKnock)[1] + dim(upInKnock)[1])

  assign(paste(deparse(substitute(knockdown)),"commonDOWN", sep = ""),
commonDownInKnock, .GlobalEnv)
  assign(paste(deparse(substitute(knockdown)),"commonUP", sep = ""),
commonUpInKnock, .GlobalEnv)
  assign(paste(deparse(substitute(knockdown)),"commonAll", sep = ""),
commonAllInKnock, .GlobalEnv)
  return(cat("Total common up/down
genes:",dim(commonUpInKnock)[1],"/",dim(commonDownInKnock)[1],'(',sum(dim(co
mmonUpInKnock)[1],dim(commonDownInKnock)[1]),')',"\n",
```

```
              "Total up+down genes in",name1,":",(dim(downInKnock)[1] +
dim(upInKnock)[1]),"\n",
              "Fraction of all",name1,"genes:",x)
          )
}
```

Appendix VI: Gene-by-gene analysis of LKB1 batches and substrate-wise contribution.

```
# APPENDIX VII

library(VennDiagram)
library(readxl)
library(RColorBrewer)
library(tidyverse)
library(hrbrthemes)
library(tm)
library(proustr)
# Prior plotting prepare excel-sheet with substrates information about
log2FC, AdjPValue and transcript identifiers

df <- read_excel(path = 'All.xlsx', sheet = 2)
df <- data.frame(df)

AMPKa1 <- df[,1]
AMPKa1 <- AMPKa1[complete.cases(AMPKa1)]
AMPKa2 <- df[,4]
AMPKa2 <- AMPKa2[complete.cases(AMPKa2)]
MARK1 <- df[,13]
MARK1 <- MARK1[complete.cases(MARK1)]
MARK2 <- df[,16]
MARK2 <- MARK2[complete.cases(MARK2)]
MARK3 <- df[,19]
MARK3 <- MARK3[complete.cases(MARK3)]
MARK4 <- df[,22]
MARK4 <- MARK4[complete.cases(MARK4)]
NUAK1 <- df[,7]
NUAK1 <- NUAK1[complete.cases(NUAK1)]
NUAK2 <- df[,10]
NUAK2 <- NUAK2[complete.cases(NUAK2)]
SIK1 <- df[,25]
SIK1 <- SIK1[complete.cases(SIK1)]
SIK2 <- df[,28]
SIK2 <- SIK2[complete.cases(SIK2)]
SIK3 <- df[,31]
SIK3 <- SIK3[complete.cases(SIK3)]
SNRK <- df[,34]
SNRK <- SNRK[complete.cases(SNRK)]

venn.diagram(
  x = list(AMPKa1, AMPKa2),
  category.names = c("AMPKa1" , "AMPKa2"),
  filename = 'AMPK_2_venn_diagramm.png',
  output=FALSE,

  imagetype="png" ,
  height = 1200 ,
  width = 1200 ,
  resolution = 500,
  compression = "lzw",
  lwd = 1.5,
  col = c('#404040', '#404040'),
  fill = c(alpha("#BF9650",0.9), alpha('#71839A',0.9)),
  cex = 0.8,
  fontfamily = "sans",
  cat.cex = 1,
  cat.fontface = "bold",
```

```
  cat.default.pos = "outer",
  cat.pos = c(0, 0),
  cat.dist = c(0.055, 0.055),
  cat.fontfamily = "sans",
  cat.col = c("#404040", '#404040')
)

venn.diagram(
  x = list(NUAK1, NUAK2),
  category.names = c("NUAK1" , "NUAK2"),
  filename = 'NUAK_venn_diagramm.png',
  output=FALSE,

  imagetype="png" ,
  height = 1200 ,
  width = 1200 ,
  resolution = 500,
  compression = "lzw",
  lwd = 1.5,
  col = c('#404040', '#404040'),
  fill = c(alpha("#BF9650",0.9), alpha('#71839A',0.9)),
  cex = 0.8,
  fontfamily = "sans",
  cat.cex = 1,
  cat.fontface = "bold",
  cat.default.pos = "outer",
  cat.pos = c(0, 0),
  cat.dist = c(0.055, 0.055),
  cat.fontfamily = "sans",
  cat.col = c("#404040", '#404040')
)

venn.diagram(
  x = list(MARK1, MARK2, MARK3, MARK4),
  category.names = c("MARK1" , "MARK2", "MARK3", "MARK4"),
  filename = 'MARK_venn_diagramm.png',
  output=FALSE,
  imagetype="png" ,
  height = 1600 ,
  width = 1800 ,
  resolution = 500,
  compression = "lzw",
  lwd = 1,
  col = c('#404040', '#404040','#404040', '#404040'),
  fill = c(alpha("#BF9650",0.9), alpha('#71839A',0.9), alpha('#6A5E40',0.9),
alpha('#AA6E29',0.9)),
  cex = 0.8,
  fontfamily = "sans",
  cat.cex = 1,
  cat.fontface = "bold",
  cat.default.pos = "outer",
  cat.pos = c(-10, 10, 0, 0),
  cat.dist = c(0.23, 0.23,0.12, 0.12),
  cat.fontfamily = "sans",
  cat.col = c("#404040", '#404040','#404040', '#404040')
)

venn.diagram(
```

```
  x = list(SIK1, SIK2, SIK3),
  category.names = c("SIK1" , "SIK2", "SIK3"),

  filename = 'SIK_venn_diagramm.png',
  output=FALSE,

  imagetype="png" ,
  height = 1200 ,
  width = 1200 ,
  resolution = 500,
  compression = "lzw",
  lwd = 1,
  col = c('#404040', '#404040','#404040'),
  fill = c(alpha("#BF9650",0.9), alpha('#71839A',0.9),
alpha('#6A5E40',0.9)),
  cex = 0.8,
  fontfamily = "sans",
  cat.cex = 1,
  cat.fontface = "bold",
  cat.default.pos = "outer",
  cat.pos = c(-10, 10, 0),
  cat.dist = c(0.05, 0.05,-0.45),
  cat.fontfamily = "sans",
  cat.col = c('#404040','#404040', '#404040')
)
```

Appendix VII: Substrate Venn-diagrams.

```
# Appendix VIII:
# R-script is applied from HOMER
(http://homer.ucsd.edu/homer/basicTutorial/affymetrix.html)
# source("http://bioconductor.org/biocLite.R")
#
# biocLite('affy')
# biocLite('oligo')
# biocLite('limma')

################################################
### Microarray analsysis                       #
### Affymetrix Human Genome U133 Plus 2.0 Array #
################################################
# load the oligo library
library(oligo)
library(limma)
library(readxl)
library(biomaRt)
# Read in the CEL files in the directory
celFiles <- list.celfiles()
affyRaw <- read.celfiles(celFiles)
# You might need to install and load a package for the specific array you
are using (this example is mouse gene 2.0 ST)
# It may try to load it automatically, but may fail.  Install & load the
library manually if this happens.

#library(pd.hg.u133.plus.2)
#library(pd.hg.u133.plus.2)
#library(pd.hg.u133a)
#library(pd.hg.u133a.2)
#library(pd.ht.hg.u133.plus.pm)
#library(pd.hugene.1.0.st.v1)
#library(pd.mouse430.2)
#library(pd.huex.1.0.st.v2)

eset <- rma(affyRaw) #RMA normalization

# Finally, save the data to an output file to be used by other programs, etc
(Data will be log2 transformed and normalized)
write.exprs(eset,file="filename_data_ALL.txt")

#library(hgu133plus2.db)
#library(hgu133a.db)
#library(hgu133a2.db)
#library(mouse4302.db)
#library("mogene10sttranscriptcluster.db")
#library(huex10stprobeset.db)

# Strategy is to create data frame objects and merge them together - put
expression info into a data frame

my_frame <- data.frame(exprs(eset))

# Put annotation information in a data frame.  To get specific fields, use
packageNameSYMBOL, where the caps part names the type of data you're after
# To get a list of available annotation information, run the packagename
with () at the end, i.e. mogene20sttranscriptcluster()
```

```
# Example
Annot <- data.frame(ACCNUM=sapply(contents(hgu133plus2ACCNUM), paste,
collapse=", "),
                    SYMBOL=sapply(contents(hgu133plus2SYMBOL), paste,
collapse=", "),
                    DESC=sapply(contents(hgu133plus2GENENAME), paste,
collapse=", "))

# Merge data frames together (like a database table join)
all <- merge(Annot, my_frame, by.x=0, by.y=0, all=T)

# Write out to a file:
write.table(all,file="filename_data.ann_ALL.txt",sep="\t")

########################################################################
### DE analsysis
########################################################################
# Example analysis: GSE26910. It contains data from breast and prostate
cancer stromas.
df <- read.table(file = 'GSE26910_data.ann_ALL.txt', header = TRUE, sep =
'\t', stringsAsFactors = FALSE)

library(edgeR)

y <- DGEList(counts = df[,5:28], genes = df[,1:4])

# Design matrix
exp_design <- data.frame(row.names = colnames(df[5:28]),
                         names = colnames(df[5:28]),
                         kd = factor(c('PNF', 'PCAF','PNF', 'PCAF','PNF',
'PCAF','PNF', 'PCAF','PNF', 'PCAF','PNF', 'PCAF',
                                       'BNF', 'BCAF','BNF', 'BCAF','BNF',
'BCAF','BNF', 'BCAF','BNF', 'BCAF','BNF', 'BCAF'))
)

design <- model.matrix(~0+exp_design$kd)

rownames(design) <- colnames(y)
colnames(design) <- c('BreastTumor', 'BreastNormal', 'ProstateTumor',
'ProstateNormal')

#vwts <- voomWithQualityWeights(y, design = design, normalize.method =
"quantile")
fit <- lmFit(my_frame, design)

# This would filter in only the highest intesity probe when multiple
sequences are present for a single gene
# o <- order(fit$Amean, decreasing=TRUE) #
https://support.bioconductor.org/p/43745/
# dup <- duplicated(fit$genes$SYMBOL[o])
# fit.unique <- fit[o,][!dup,]

# Make contrasts:
my.contrasts <- makeContrasts(
  ProstateTumor = ProstateTumor - ProstateNormal,
  BreastTumor = BreastTumor - BreastNormal,
  levels = design
)
```

```
fit.de <- contrasts.fit(fit, my.contrasts)
fit.de <- eBayes(fit.de)

summary(decideTests(fit.de, adjust.method = "fdr", p.value = 0.05))

for (i in 1:2) {
  res <- topTable(fit.de, coef = i, number = Inf, adjust.method = "fdr")
  write.table(res, file = paste(colnames(my.contrasts)[i],
"vsNormal_ALL.txt", sep = ""), row.names = TRUE, sep = "\t", quote = FALSE)
}

###############################################################################
### Post-annotation
###############################################################################

df2 <- read.delim("ProstateTumorvsNormal_ALL.txt", check.names = FALSE,
stringsAsFactors = FALSE)
df2 <- read.delim("BreastTumorvsNormal_ALL.txt", check.names = FALSE,
stringsAsFactors = FALSE)

df2 <- merge(Annot, df2, by.x=0, by.y=0, all=T)
write.table(df2, file = "BreastTumorvsNormal_ALL.txt", row.names = TRUE, sep
= "\t", quote = FALSE) # Re-save the DE-dataframe with annotations to the
sane file-name.

###############################################################################
### Probe filtering
###############################################################################

df3 <- read.delim("ProstateTumorvsNormal_ALL.txt", check.names = FALSE,
stringsAsFactors = FALSE)
df3 <- read.delim("BreastTumorvsNormal_ALL.txt", check.names = FALSE,
stringsAsFactors = FALSE)

df3 <- df3[,c(1,3,5:10)]

# Remove genes named 'NA'; some probes didn't contain any annotated gene
symbol at all.
df3 <- df3[complete.cases(df3[,2]),]

unique_genes_in_df4 <- unique(df3[,2])

result <- NULL
for (i in 1:length(unique_genes_in_df4)) {
  gene <- unique_genes_in_df4[i]
  genes_in_df4 <- df3[df3[,2] %in% gene, ]
  genes_in_df4 <- genes_in_df4[order(genes_in_df4[,7]) , ]
  result <- rbind(result, genes_in_df4[1,])
}

write.table(x = result, file =
'ProstateTumorvsNormal_probefiltered_ALL.txt', row.names = F, sep = "\t",
quote = FALSE)
write.table(x = result, file = 'BreastTumorvsNormal_probefiltered_ALL.txt',
row.names = F, sep = "\t", quote = FALSE)

###############################################################################
```

```
### GSEA file formatting:
### *.rnk.txt
### *.500_ALL.gmx
##############################################################################

df4 <- read.table(file = 'ProstateTumorvsNormal_probefiltered_ALL.txt',
header = TRUE, sep = '\t', stringsAsFactors = FALSE)
df4 <- read.table(file = 'BreastTumorvsNormal_probefiltered_ALL.txt', header
= TRUE, sep = '\t', stringsAsFactors = FALSE)


# GSEA-file
GSEA_file <- df4[order(df4[,3]) , ]
GSEA_file <- GSEA_file[,c(2,3)]

write.table(x = GSEA_file, file =
'GSE26910_ProstateTumorvsNormal_GSEA_Hs.rnk.txt', row.names = FALSE, sep =
"\t", quote = FALSE)
write.table(x = GSEA_file, file =
'GSE26910_BreastTumorvsNormal_GSEA_Hs.rnk.txt', row.names = FALSE, sep =
"\t", quote = FALSE)

GSEA_file <- read.table(file =
'GSE26910_BreastTumorvsNormal_GSEA_Hs.rnk.txt', header = T, sep = '\t',
stringsAsFactors = F)

result_signf_up <-  GSEA_file[1:250,]# Total 250
result_signf_down <- GSEA_file[((dim(GSEA_file)[1])-
250+1):(dim(GSEA_file)[1]),] # Total 250
result_signf_1 <- rbind(result_signf_up, result_signf_down)

result_signf_up <-  GSEA_file[1:500,]# Total 250
#result_signf_down <- GSEA_file[15684:15933,] # Total 250
result_signf_2 <- rbind(result_signf_up)

#result_signf_up <-  GSEA_file[1:250,]# Total 250
result_signf_down <- GSEA_file[((dim(GSEA_file)[1])-
500+1):(dim(GSEA_file)[1]),] # Total 250
result_signf_3 <- rbind(result_signf_down)

result_GSEA_names_1 <- result_signf_1[,c(1)]
result_GSEA_names_2 <- result_signf_2[,c(1)]
result_GSEA_names_3 <- result_signf_3[,c(1)]

write.table(x = result_GSEA_names_1, file =
'GSE26910_BreastTumorvsNormal_500_ALL.gmx', row.names = FALSE, sep = "\t",
quote = FALSE)
write.table(x = result_GSEA_names_2, file =
'GSE26910_BreastTumorvsNormal_500_DOWN_ALL.gmx', row.names = FALSE, sep =
"\t", quote = FALSE)
write.table(x = result_GSEA_names_3, file =
'GSE26910_BreastTumorvsNormal_500_UP_ALL.gmx', row.names = FALSE, sep =
"\t", quote = FALSE)
```

Appendix VIII: Example R-script for microarray data analysis. Example analysis of GSE26910. Script contains filtering, background correction, annotating, DE-analysis and GSEA-file formatting (ranked gene list and (Top-/Bottom- and overall-) gene set).

```
# Appendix IX:
# Input the GEO2R generated differential expression table and filter
ambiguous expressions:
df <- read.table(file = 'filename.txt', header = TRUE, sep = '\t',
stringsAsFactors = FALSE)
df <- df[,c(8,6,3,2)]
df[,1] <- gsub("/.*","",df[,1])

#df$'refseq_mRNA' <- sub("\\..*", "", df[,1])
#df <- df[,c(5,2:4)]
df <- df[complete.cases(df[,1]) , ]
df <- df[!(df[,1]==""), ]

# Include the array-probes with the highest statisticl significance:
unique_genes_in_df4 <- unique(df[,1])

result <- NULL
for (i in 1:length(unique_genes_in_df4)) {
  gene <- unique_genes_in_df4[i]
  genes_in_df4 <- df[df[,1] %in% gene, ]
  genes_in_df4 <- genes_in_df4[order(genes_in_df4[,3]) , ]
  result <- rbind(result, genes_in_df4[1,])
}

#-------------------------------------------------------------------------
# biomaRt homolog fetching; the script depends on what kind of identifiers
the original arrya was using.

library(biomaRt)
# ensembl = useMart("ensembl")
# datasets = listDatasets(ensembl)
# genes = result$GB_ACC
#
# human = useMart("ensembl", dataset = "hsapiens_gene_ensembl")
# mouse = useMart("ensembl", dataset = "mmusculus_gene_ensembl")
# G_list <- getLDS(attributes = c('mgi_symbol','refseq_mrna'),
#                  filters = "refseq_mrna", values = genes,mart = mouse,
#                  attributesL = c("hgnc_symbol"), martL = human)
#
# mergedData <- merge(x = df,
#                     y = G_list,
#                     by.x = "GB_ACC",
#                     by.y = "RefSeq.mRNA.ID"
# )

#--
#-----
#--

ensembl = useDataset("hsapiens_gene_ensembl", useMart("ensembl"))

#result[,1] <- gsub("[.].*$", "", result[,1])
genes = result[,1]

G_list <- getBM(attributes= c('hgnc_symbol','refseq_mrna'),
                filters= 'refseq_mrna',
                values=genes,
                mart= ensembl)
```

```
mergedData <- merge(x = result,
                    y = G_list,
                    by.x = 'GB_ACC',
                    by.y = "refseq_mrna"
)
# # #-------------------------------------------------------------------------
mergedData <- mergedData[,c(5,2:4)]
mergedData <- unique(mergedData)
mergedData <- mergedData[!(mergedData$hgnc_symbol==""), ]
mergedData <- mergedData[complete.cases(mergedData) , ]
# Re-do filtering
unique_genes_in_df4 <- unique(mergedData[,1])
result <- NULL
for (i in 1:length(unique_genes_in_df4)) {
  gene <- unique_genes_in_df4[i]
  genes_in_df4 <- mergedData[mergedData[,1] %in% gene, ]
  genes_in_df4 <- genes_in_df4[order(genes_in_df4[,3]) , ]
  result <- rbind(result, genes_in_df4[1,])
}
write.table(x = result, file = 'filename_probefiltered_ALL.txt', row.names =
F, sep = "\t", quote = FALSE)

#-------------------------------------------------------------------------
# GSEA file formatting:

df2 <- read.table(file = 'filename_probefiltered_ALL.txt', header = TRUE,
sep = '\t', stringsAsFactors = FALSE)
# GSEA-file
GSEA_file <- df2[order(df2[,2]) , ]
GSEA_file <- GSEA_file[,c(1,2)]
# Optional filtering
# GSEA_file_new <- gsub("/.*","",GSEA_file[,1])
# GSEA_file_new <- gsub("//.*","",GSEA_file_new)
# GSEA_file_new <- gsub("///.*","",GSEA_file_new)
# GSEA_file_new <- gsub("////.*","",GSEA_file_new)
# GSEA_file_new <- gsub("@.*","",GSEA_file_new)
# GSEA_file[,3] <- GSEA_file_new
# GSEA_file <- GSEA_file[,c(3,2)]
# GSEA_file_new <- GSEA_file[ grep("//", GSEA_file$Gene.symbol, invert =
TRUE) , ]
write.table(x = GSEA_file, file = 'GSE90505_CAF-vs-CNTRL_GSEA_Hs.rnk.txt',
row.names = FALSE, sep = "\t", quote = FALSE)

# Gene set file generation: bottom, down
GSEA_file <- read.table(file = 'filename_GSEA_Hs.rnk.txt', header = T, sep =
'\t', stringsAsFactors = F)

result_signf_up <-  GSEA_file[1:250,]# Total 250
result_signf_down <- GSEA_file[18753:19002,] # Total 250
result_signf_1 <- rbind(result_signf_up, result_signf_down)

result_signf_up <-  GSEA_file[1:500,]# Total 250
#result_signf_down <- GSEA_file[18503:19002,] # Total 250
result_signf_2 <- rbind(result_signf_up)

#result_signf_up <-  GSEA_file[1:250,]# Total 250
result_signf_down <- GSEA_file[18503:19002,] # Total 250
```

```
result_signf_3 <- rbind(result_signf_down)

result_GSEA_names_1 <- result_signf_1[,c(1)]
result_GSEA_names_2 <- result_signf_2[,c(1)]
result_GSEA_names_3 <- result_signf_3[,c(1)]

write.table(x = result_GSEA_names_1, file = 'filename_500_ALL.gmx',
row.names = FALSE, sep = "\t", quote = FALSE)
write.table(x = result_GSEA_names_2, file = 'filename__500_DOWN_ALL.gmx',
row.names = FALSE, sep = "\t", quote = FALSE)
write.table(x = result_GSEA_names_3, file = 'filename__500_UP_ALL.gmx',
row.names = FALSE, sep = "\t", quote = FALSE)
```

Appendix IX: Example R-script for GEO2R DE-table analysis. Here, the differential expression is conducted inside the browser-based analysis tool (Gene Expression Omnibus) and the expression table is saved and imported into the R-environment for further formatting. The script depends on the also on the input file (array type, species, identifiers etc.).

```
# APPENDIX X
library(gplots)
library(pheatmap)
library(readxl)
library(ggplot2)
library(DESeq2)
library(data.table)
library(pheatmap); library("RColorBrewer")
library(cluster)
library(dplyr)

####################################
###   Import file and            ###
###   generate common dataframe  ###
####################################

file_list <- list.files()

# Format the files to fit together
for (file in file_list){

  temp_dataset <-read.table(file, header=TRUE, sep="\t"
                            , col.names = c('ID',as.character(file))
                            )
  temp_dataset[,1] <- gsub("/.*","",temp_dataset[,1])
  temp_dataset[,1] <- gsub(";.*$", "", temp_dataset[,1])
  temp_dataset[,1] <- gsub("_.*$", "", temp_dataset[,1])

  if (length(unique(temp_dataset[,1])) == dim(temp_dataset)[1]) {
    row.names(temp_dataset)<-temp_dataset[,1]
    temp_dataset<-data.frame(temp_dataset)
    assign(paste(file), temp_dataset, .GlobalEnv)
    rm(temp_dataset)

  } else {
    unique_genes_in_df4 <- unique(temp_dataset[,1])
    result <- NULL
    for (i in 1:length(unique_genes_in_df4)) {
      gene <- unique_genes_in_df4[i]
      genes_in_df4 <- temp_dataset[temp_dataset[,1] %in% gene, ]
      genes_in_df4 <- genes_in_df4[order(genes_in_df4[,2]) , ]
      result <- rbind(result, genes_in_df4[1,])
    }

    temp_dataset<-result
    row.names(temp_dataset)<-temp_dataset[,1]
    temp_dataset<-data.frame(temp_dataset)
    assign(paste(file), temp_dataset, .GlobalEnv)
    rm(temp_dataset)
  }
}

flightsList<-list(GSE101665_CAFvsNF_GSEA.Hs.rnk.txt, `GSE114056_CAF-vs-
CNTRL_GSEA_Hs.rnk.txt`, `GSE116167_44As3-vs-NF_GSEA_Hs.rnk.txt`,
                  `GSE116167_HSC44PE-vs-NF_GSEA_Hs.rnk.txt`,`GSE116679_HIGH-
vs-LOW_GSEA_Hs.rnk.txt`, GSE118624_GSEA.rnk.txt,
                 GSE1724_FASScvsCNTRL_GSEA_Hs.rnk.txt,
GSE1724_IPFvsCNTRL_GSEA_Hs.rnk.txt, GSE20086_CAFvsNF_GSEA_Hs.rnk.txt,
```

```
                  GSE22862_CAFvsNF_GSEA_Hs.rnk.txt,
                  GSE29270_CAFvsNF_GSEA_Hs.rnk.txt,
GSE34312_CancerVSnormal_GSEA_Hs.rnk.txt, `GSE35364_CAF-vs-
CNTRL_GSEA_Hs.rnk.txt`,
                  `GSE35364_coCAF-vs-
CNTRL_GSEA_Hs.rnk.txt`,GSE37738_cSCCvsNF_GSEA_Hs.rnk.txt,
GSE37738_RDEBcSCCvsNF_GSEA_Hs.rnk.txt,
                  GSE37738_RDEBvsNF_GSEA_Hs.rnk.txt,
GSE38517_CAF_MLvsNF_GSEA_Hs.rnk.txt, GSE38517_CAFvsNF_GSEA_Hs.rnk.txt,
                  GSE38517_DysplasticvsNF_GSEA_Hs.rnk.txt,
GSE40839_IntPneumvsNF_GSEA_Hs.rnk.txt,
GSE40839_ScleroCAFvsNF_GSEA_Hs.rnk.txt,
                  GSE43770_CAFvsNF.Hs.rnk.txt, `GSE44723_rapidIPF-vs-
CNTRL_GSEA_Hs.rnk.txt`, `GSE44723_stableIPF-vs-CNTRL_GSEA_Hs.rnk.txt`,
                  `GSE45256_ADENO-vs-NF_GSEA_Hs.rnk.txt`, `GSE45256_CAF-vs-
NF_GSEA_Hs.rnk.txt`, `GSE45256_HYP-vs-NF_GSEA_Hs.rnk.txt`,
                  `GSE45686_IPF-vs-CNTRL_GSEA_Hs.rnk.txt`,
GSE46824_GSEA.rnk.txt, `GSE48397_CAF-vs-CNTRL_GSEA_Hs.rnk.txt`,
`GSE62740_CAF-vs-CNTRL_GSEA_Hs.rnk.txt`,
                  GSE67250_GSEA_Hs.rnk.txt,
GSE68164_CAFvsNF_GSEA_Hs.rnk.txt, GSE70468_CAFvsNF_GSEA_Hs.rnk.txt,
                  GSE71078_CAFvsNF_GSEA_Hs.rnk.txt, `GSE71351_RAPID_IPF-vs-
CNTRL_GSEA_Hs.rnk.txt`, `GSE71351_SLOW_IPF-vs-CNTRL_GSEA_Hs.rnk.txt`,
                  GSE73728_day11_TCLNvsPBS_GSEA_Hs.rnk.txt,
GSE73728_day4_TCLNvsPBS_GSEA_Hs.rnk.txt, GSE79786_PAHvsCNTL_GSEA_Hs.rnk.txt,
                  GSE83314_MUTvsWT_GSEA.Hs.rnk.txt,
GSE83834_MUTvsWT_GSEA.Hs.rnk.txt, `GSE85606_CAF-vs-CNTRL_GSEA.Hs.rnk.txt`,
                  `GSE86256_CAF-vs-CNTRL_GSEA_Hs.rnk.txt`,
GSE90607_INFLvsNF_GSEA_Hs.rnk.txt, GSE90607_STENOvsNF_GSEA_Hs.rnk.txt,
                  `GSE93313_TRANSWELL-vs-3D_GSEA.Hs.rnk.txt`,
`GSE99816_INFL-vs-CNTRL_GSEA_Hs.rnk.txt`, `GSE99816_NON_INFL-vs-
CNTRL_GSEA_Hs.rnk.txt`,
                  `GSE99816_STENO-vs-CNTRL_GSEA_Hs.rnk.txt`)

# Merge the DE dataframes
df<-Reduce(function(x, y) merge(x, y, all=TRUE), flightsList)

df1<-df
#df <- df1
df <- df[-c(1) , ]

# Filter out genes which are out in majority of arrays/sequencings
na_count <- apply(df, 1, function(x) sum(is.na(x)))

# Taking into account that not all the transcripts are identified in all of
the samples.
# If a gene is expressed in more the 90% of samples, it is accepted; sample
that do not
# express the gene in this 90% will be denoted as zeroes (NA values are not
accepted
# in hierarchical clustering).
rows<-NULL
for (i in 1:dim(df)[1]) {
  if (na_count[i] <= 5) {
    df[i,][is.na(df[i,] ) ] <- 0
    rows<-rbind(rows, df[i,1])
  }
}
```

```
df2 <- df[df$ID %in% rows, ]
row.names(df2)<-df2[,1]
#write.table(x = df2, file = 'df2_IndirectCAF.txt', row.names = FALSE, sep =
"\t", quote = FALSE)
df2<-df2[,-c(1)]

# (Convert to Z-scores???)
commonDF<-df2
commonDF<-data.matrix(commonDF)
commonDF<-commonDF[order(rowSums(commonDF),decreasing=T),]
# Only include to TOP/BOTTOM1000 expressed genes
commonDFup <-  commonDF[1:1000,]
commonDFdown <- commonDF[7654:8653,]
commonDF <- rbind(commonDFup, commonDFdown)

#-------------------------------------------------------------------------
#########################################################################
###    k-means clusteirng
###    http://girke.bioinformatics.ucr.edu/GEN242/mydoc_Rclustering_4.html
#########################################################################

yscaled <- t(scale(t(commonDF)))
apply(yscaled, 1, sd)

dist(commonDF[1:4,], method = "euclidean")

# Correlation matrix
c <- cor(t(commonDF), method="pearson")
#as.matrix(c)[1:4,1:4]

# Correlation-based distance matrix
d <- as.dist(1-c)
#as.matrix(d)[1:4,1:4]

#------------------------
# Hierarchical clustering

hr <- hclust(d, method = "complete", members=NULL)
names(hr)

#par(mfrow = c(1, 2)); plot(hr, hang = 0.1); plot(hr, hang = -1)

hc <- hclust(as.dist(1-cor(yscaled, method="spearman")), method="complete")
#mycol <- colorpanel(40, "darkblue", "yellow", "white")

colfunc <- colorRampPalette(c("#d36000", "white", "#00306e"))

#----------------------------
# K-Means Clustering with PAM

pamy <- pam(d, 4)
(kmcol <- pamy$clustering)

heatmap.2(commonDF,
          Rowv=as.dendrogram(hr),
          Colv=as.dendrogram(hc),
          #col = brewer.pal(9,"PuOr"),
```

```
          col = colfunc(45),
          scale="row",
          density.info="none",
          trace="none",
          #RowSideColors=as.character(kmcol),
          labRow = F,
          margins = c(12, 1),
          cexCol = 0.95,
          srtCol = 35,
          lhei = c(0.6,5)
)
```

Appendix X: R-script for hierarchical cluster of published CAF-datasets (Fig3B). As an example the indirect patient samples. Hierarchical clustering was performed similarly in MEF-samples (Fig2A), direct patient samples (Fig3A) and with curated CAF-genes (FigS4).

```
# APPENDIX XI
library(gplots)
library(pheatmap)
library(readxl)
library(ggplot2)
library(DESeq2)
library(data.table)
library(pheatmap); library("RColorBrewer")
library(cluster)
library(dplyr)

####################################
###   Import file and            ###
###   generate common dataframe  ###
####################################

inFilePaths = list.files(path=".", pattern=glob2rx("*"), full.names=TRUE)

for (inFilePath in inFilePaths) {
  inFileData = read.csv(inFilePath)
  print(inFilePath)
}
#-------------------------------------------------------------------------
# Apparently there are still some dublicated rows. These are not allowed
when merging the dataframes, thus need to be removed
# (internal for-loop)
file_list <- list.files()

for (file in file_list){

  temp_dataset <-read.table(file, header=TRUE, sep="\t", col.names =
c('ID',as.character(file)))
  temp_dataset[,1] <- gsub("/.*","",temp_dataset[,1])
  temp_dataset[,1] <- gsub(";.*$", "", temp_dataset[,1])
  temp_dataset[,1] <- gsub("_.*$", "", temp_dataset[,1])

  if (length(unique(temp_dataset[,1])) == dim(temp_dataset)[1]) {
    row.names(temp_dataset)<-temp_dataset[,1]
    temp_dataset<-data.frame(temp_dataset)
    assign(paste(file), temp_dataset, .GlobalEnv)
    rm(temp_dataset)

  } else {
    unique_genes_in_df4 <- unique(temp_dataset[,1])
    result <- NULL
    for (i in 1:length(unique_genes_in_df4)) {
      gene <- unique_genes_in_df4[i]
      genes_in_df4 <- temp_dataset[temp_dataset[,1] %in% gene, ]
      genes_in_df4 <- genes_in_df4[order(genes_in_df4[,2]) , ]
      result <- rbind(result, genes_in_df4[1,])
    }

    temp_dataset<-result
    row.names(temp_dataset)<-temp_dataset[,1]
    temp_dataset<-data.frame(temp_dataset)
    assign(paste(file), temp_dataset, .GlobalEnv)
    rm(temp_dataset)
  }
```

```
}

flightsList<-list(GSE101665_CAFvsNF_GSEA.Hs.rnk.txt, `GSE114056_CAF-vs-
CNTRL_GSEA_Hs.rnk.txt`, `GSE116167_44As3-vs-NF_GSEA_Hs.rnk.txt`,
                  `GSE116167_HSC44PE-vs-NF_GSEA_Hs.rnk.txt`,
GSE118624_GSEA.rnk.txt, GSE20086_CAFvsNF_GSEA_Hs.rnk.txt,
                  `GSE35364_CAF-vs-CNTRL_GSEA_Hs.rnk.txt`,
GSE37738_cSCCvsNF_GSEA_Hs.rnk.txt, GSE37738_RDEBcSCCvsNF_GSEA_Hs.rnk.txt,
                  GSE37738_RDEBvsNF_GSEA_Hs.rnk.txt,
GSE38517_CAF_MLvsNF_GSEA_Hs.rnk.txt, GSE40839_IntPneumvsNF_GSEA_Hs.rnk.txt,
                  GSE40839_ScleroCAFvsNF_GSEA_Hs.rnk.txt,
`GSE44723_rapidIPF-vs-CNTRL_GSEA_Hs.rnk.txt`, `GSE44723_stableIPF-vs-
CNTRL_GSEA_Hs.rnk.txt`,
                  GSE46824_GSEA.rnk.txt, GSE70468_CAFvsNF_GSEA_Hs.rnk.txt,
GSE83834_MUTvsWT_GSEA.Hs.rnk.txt,
                  `GSE99816_INFL-vs-CNTRL_GSEA_Hs.rnk.txt`,
`GSE99816_NON_INFL-vs-CNTRL_GSEA_Hs.rnk.txt`, `GSE99816_STENO-vs-
CNTRL_GSEA_Hs.rnk.txt`)

# Merge the DE dataframes
df<-Reduce(function(x, y) merge(x, y, all=TRUE), flightsList)

df1<-df
df <- df[-c(1) , ]

# Filter out genes which are out in majority of arrays/sequencings
na_count <- apply(df, 1, function(x) sum(is.na(x)))

rows<-NULL
for (i in 1:dim(df)[1]) {
  if (na_count[i] <= 2) {
    df[i,][is.na(df[i,] ) ] <- 0
    rows<-rbind(rows, df[i,1])
  }
}

#df$na_count_2 <- apply(df, 1, function(x) sum(is.na(x)))

df2 <- df[df$ID %in% rows, ]
row.names(df2)<-df2[,1]
#write.table(x = df2, file =
'CommonGenes_Cluster1_PatientDerivedDirect.txt', row.names = FALSE, sep =
"\t", quote = FALSE)
df2<-df2[,-c(1)]

# Z-scoring, though his didn't seem to result anything meaningfull since
normalized values sum up to zero. Extremely small values etc. not good for
ranking the genes.
# df2_scaled <- scale(t(df2))
# df2_scaled <- t(df2_scaled)
# df2_scaled <- data.frame(df2_scaled)
# df2_scaled$'SUM'<-rowSums(df2_scaled)
# df2$'Zscores' <- rowSums(df2_scaled)/13

df2$'Average'<-rowSums(df2)/21
df2$'ID'<-row.names(df2)

df3<-df2[,c(23,22)]
```

```
df3<-df3[order(df3[,2]) , ]
write.table(x = df3, file = 'CommonGenes_Cluster1_2.0_NEW_Hs.GSEA.rnk.txt',
row.names = FALSE, sep = "\t", quote = FALSE)
df3<-read.table(file = 'CommonGenes_Cluster1_2.0_NEW_Hs.GSEA.rnk.txt',
header = T, sep = '\t', stringsAsFactors = F)

df4UP <-  df3[1:250,]
write.table(x = df4UP, file =
'PatientDerivedIndirect_2.0_NEW_CommonGenes_Cluster1_DOWN_Hs.GSEA.rnk.txt',
row.names = FALSE, sep = "\t", quote = FALSE)
df4DOWN <- df3[10014:10263,]
write.table(x = df4DOWN, file =
'PatientDerivedIndirect_2.0_NEW_CommonGenes_Cluster1_UP_Hs.GSEA.rnk.txt',
row.names = FALSE, sep = "\t", quote = FALSE)
```

Appendix XI: Merging and GSEA-formatting of patient CAF-clusters. After identification of patient clusters with Appendix X, the individual CAF-data and dinned according to hierarchical clustering. As in Appendix X, cluster-specific data are combined and merged (averaged). Finally ranked GSEA-gene list and Top-/Bottom-gene sets are generated for following enrichment analyses. As an example is patient derived, indirect cluster II.